

Review

Not peer-reviewed version

Computer Vision-Based Gait Recognition on the Edge: A Survey on Feature Representations, Models, and Architectures

[Edwin Salcedo](#) *

Posted Date: 11 October 2024

doi: 10.20944/preprints202410.0903.v1

Keywords: Gait recognition; gait analysis; edge computing; computer vision; edge ai



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Review

Computer Vision-Based Gait Recognition on the Edge: A Survey on Feature Representations, Models, and Architectures

Edwin Salcedo

Department of Mechatronics Engineering, Universidad Catolica Boliviana "San Pablo", La Paz 4807, Bolivia;
esalcedo@ucb.edu.bo

Abstract: Computer vision-based gait recognition (CVGR) is a biometric technology that has gained considerable attention in recent years due to its non-invasive, unobtrusive, and difficult-to-conceal nature. Current CVGR systems often transmit collected data to a cloud server for machine-learning-based gait pattern recognition. While effective, this cloud-centric approach can lead to increased system response times. Alternatively, the emerging paradigm of edge computing, which involves moving computational processes to local devices, offers the potential to reduce latency, enable real-time surveillance, and eliminate reliance on internet connectivity. Furthermore, recent advancements in low-cost, compact microcomputers capable of handling complex inference tasks (e.g., Jetson Nano, Xavier NX, AGX Xavier, Raspberry Pi, and Khadas) have created exciting opportunities for deploying CVGR systems at the edge. This paper reports the state of the art in gait data acquisition modalities, feature representations, models, and architectures for CVGR systems amenable for edge computing. Moreover, this paper addresses the general limitations and highlights new avenues for future research in the promising intersection of CVGR and edge computing.

Keywords: Gait recognition; gait analysis; edge computing; computer vision; edge ai

1. Introduction

In an era marked by the proliferation of smart devices and surveillance systems, biometrics has emerged as a crucial field for secure and convenient identity verification. Among various biometric modalities, gait recognition, which analyses an individual's unique walking pattern, provides a non-invasive and potentially unobtrusive identification method [1]. However, traditional gait recognition systems, often reliant on centralized cloud-based architectures, face challenges in terms of latency, bandwidth consumption, and privacy concerns. The emergence of edge computing, which brings computation and data storage closer to the source, presents a compelling paradigm shift for gait recognition systems. This approach unlocks numerous opportunities not only for biometrics and security, but also for healthcare, rehabilitation, and human-computer interaction projects.

Gait data can be gathered using various wearable sensors, including accelerometers, gyroscopes, and force/pressure sensors [2]. In contrast, computer vision-based gait recognition (CVGR) systems primarily rely on 2D or 3D imaging sensors to capture gait data from long distances, even without the subject's active participation [1]. CVGR systems offer several advantages: they are non-invasive, robust to impersonation, function well over long distances, do not need high-quality cameras, and can be combined with other sensing modalities [3]. As illustrated in Figure 1a, recent years have witnessed a shift in gait pattern recognition, moving away from traditional approaches towards modern techniques based on deep learning (DL). These DL techniques train deep neural networks on high-dimensional gait data for pattern recognition. The high accuracy of deep neural networks in gait identification has made them the preferred approach, creating new opportunities for deploying CVGR systems in real-world environments.

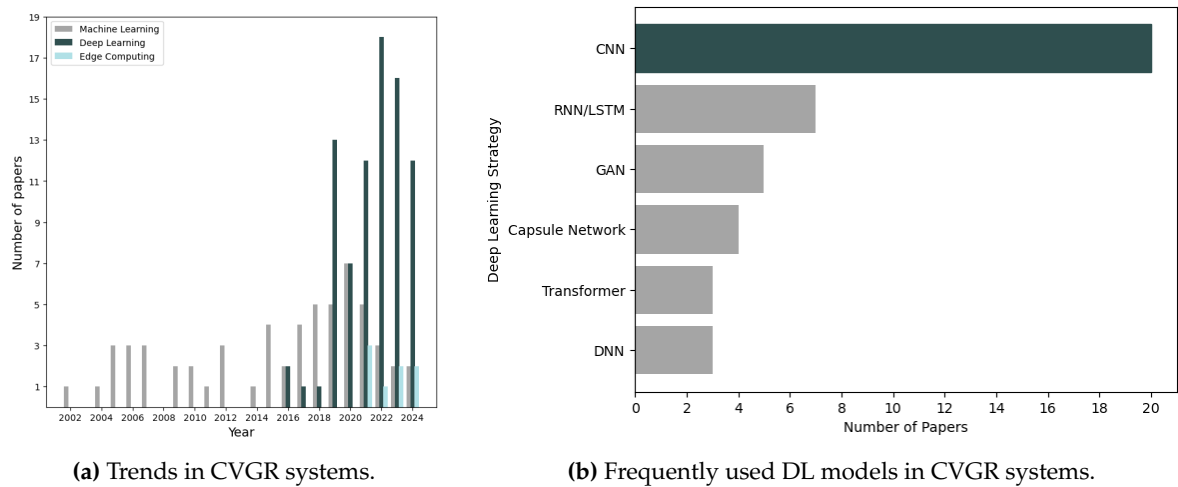


Figure 1. A comparative analysis illustrating the growing preference for DL architectures. The illustrations summarizes the findings from the papers reviewed in this survey.

Despite advancements, challenges such as varying view angles, intra-class variations, body segment localisation, and occlusions persist in CVGR systems. Researchers have recently addressed some of these challenges by increasing the complexity of the architectures [4–7]. However, such complexity comes at the expense of high computational costs and resource requirements, hindering on-site deployment and necessitating resource-intensive environments. Fortunately, ongoing advancements in hardware and software for embedded systems are leading to smaller, more powerful computers (e.g., Jetson Nano, Xavier NX, AGX Xavier, Raspberry Pi, Khadas VIM, and Google Coral Edge TPU), enabling the embedding of DL models even in resource-constrained settings. This trend positions edge computing as a promising solution for deploying DL-based gait recognition in applications like intelligent, real-time surveillance and smart authentication systems.

Several comprehensive reviews have been conducted on CVGR systems, focusing on DL [8], appearance-based gait recognition [9], and energy image-based gait recognition [10]. However, to the best of the authors’ knowledge, no comprehensive review has specifically addressed the intersection between gait recognition systems based on computer vision and edge computing. This gap in the literature highlights the value of the present work. Therefore, we conduct a thorough survey of CVGR systems, with a focus on 2D imaging-based gait recognition due to its suitability for low-resource end devices. We will also examine the benefits and opportunities associated with deploying CVGR on edge devices. The paper also examines the benefits and opportunities associated with deploying CVGR on edge devices, as well as the obstacles to such deployment, considering factors such as computational efficiency, real-time performance, and data security.

The structure of this survey is illustrated in Figure 2. Firstly, Section 2 details common sensors and methods for acquiring gait data. Following this, Section 3 explores various techniques for representing and processing gait features extracted from 2D images in CVGR systems. We then delve into different dimensionality reduction techniques aimed at achieving computational efficiency in edge computing architectures. Section 5 then provides a detailed overview of both traditional and modern models for gait feature classification. Next, Section 6 examines various edge computing architectures, while Section 7 presents techniques to accelerate machine learning inference. The convergence of gait recognition and edge computing is discussed in Section 8. Finally, the paper conclude by discussing and summarising our findings in Sections 9 and 10, respectively.

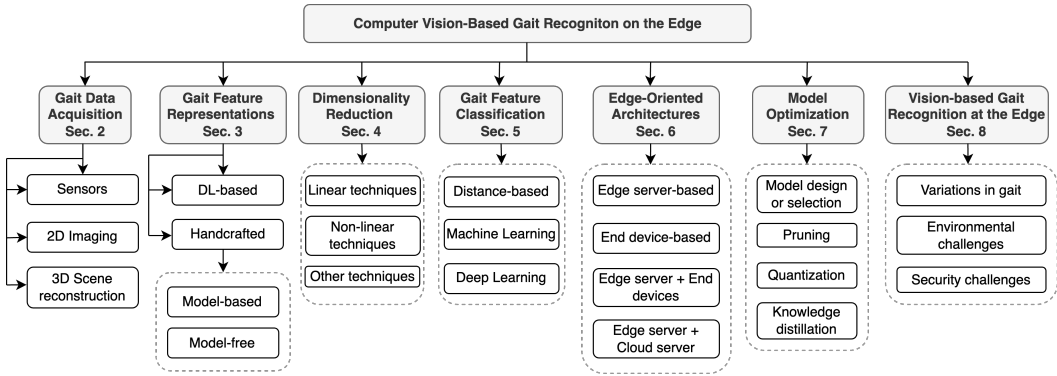


Figure 2. General structure of this survey paper and our proposed taxonomy of the existing technologies that facilitate on-device deployment of CVGR systems for real-time gait recognition.

2. Data Acquisition Methods for Gait Recognition

2.1. Sensor-Based Methods

Gait recognition systems primarily utilise two data collection methods: sensors and 2D imaging. However, a growing body of research is exploring alternative gait data acquisition approaches, including depth sensing, stereo vision, point clouds, and optical flow [8]. Regarding the sensor-based approach, floor sensors are commonly employed to register the pressure exerted by walking individuals, converting this data into signals [11,12]. These signals are then processed to either distinguish unique gait characteristics among individuals or identify gait anomalies. The main disadvantage of floor sensors, however, is their expensive deployment, which often necessitates controlled indoor environments. Similarly, wearable sensors, such as accelerometers, gyroscopes, goniometers, and electromyography, have also been applied to gait analysis [2]. These sensors can be connected (wirelessly or wiredly) to embedded devices, smartphones, or other wearables, aiming to capture dynamic parameters of the human body for gait analysis [2]. For instance, researchers in [13] recently explored edge computing for gait recognition using gyroscope data. They deployed a 4-layer CNN on an Arduino Nano 33 BLE, a smartphone, and a Brainchip Akida processor, achieving real-time person identification. Despite this progress, the present review does not discuss works developed under this sensing modality, given that they require subject cooperation.

2.2. 2D Imaging-Based Methods

CVGR systems based on 2D imaging utilise 2D image or video data of human gait sequences, typically captured by a digital camera, and analyse this data using 2D image or video processing techniques. This sensing modality has been widely researched, leading to the creation of several well-established datasets for development and performance evaluation, some of the most representative of which are listed in Table 1. The choice of specific image processing techniques after data acquisition is significantly influenced by the design of the target CVGR system. Common processing methods include:

- Noise reduction to minimise the impact of noise or artifacts in the video data.
- Background subtraction and silhouette extraction to isolate the moving human subject (foreground) from the static background.
- Gait cycle detection and normalisation to identify and standardize the repetitive pattern of gait cycles.

The present review primarily focuses on this sensing approach due to the advantages of processing 2D data on resource-constrained devices, including faster computation and the ability to utilise low-cost, standard cameras.

Table 1. Datasets for 2D Imaging-based gait recognition sorted by year.

Name & Reference	Year	Subjects	Sequences	Views	Variations	Environment
CASIA-E [14]	2022	1,014	778,752	1	Dressing, Carrying	Outdoor
ReSGait [15]	2021	172	870	1	Clothing, Carrying, Trajectories	Indoor
GREW [16]	2021	26,345	128,671	882	Clothing, Carrying, Occlusion, Viewpoint, Background	Outdoor
VersatileGait [17]	2021	11,000	1,320,000	44		Unity3D
OU-MVLP * [18]	2020	10,307	268,086	14	Age, Gender, Walking Style	Indoor
KY4D * [19]	2014	42	84	16	Viewpoint	Indoor
SOTON * [20]	2011	300	5,000	12	Curve	Indoor
SAIVT-DGD [21]	2011	35	700	1	Viewpoint	Indoor
CASIA-C [22]	2006	153	1,530	1	Speed, Carrying, Shoes	Indoor
CASIA-B [23]	2006	124	13,640	11	Speed, Walking Surface	Outdoor
CASIA-A [24]	2003	20	240	1	Clothing, Carrying, Walking Surface	Indoor
UCSD [25]	1999	6	42	1	Walking Direction	Outdoor
					Clothing, Carrying	Outdoor

* The dataset comes in different versions. The specified details are from its most recent version.

2.3. 3D Scene Reconstruction-Based Methods

While 2D image sensors are the foundation of computer vision systems, other sensors can also play an important role, particularly in 3D scene reconstruction. For instance, technologies such as depth sensing, 3D lasers, and stereo vision aim to capture the complete 3D structure of real-world objects and have recently shown promising results for gait recognition. However, their current high computational demands and associated costs limit their widespread deployment in edge computing-based architectures.

2.3.1. Depth Sensing

CVGR systems have leveraged depth cameras (e.g., Microsoft Kinect, Zed, and Intel RealSense) due to their ability to provide gait information based on a skeleton model [26–29]. This approach, also known as model-based gait recognition, utilises both RGB images and depth maps to identify joint points in the human body. It then models the anatomical components to derive a unique gait signature [30]. Because they are less sensitive to lighting variations and occlusion issues, depth cameras provide a more consistent input for gait recognition. However, several studies have found this approach to be computationally expensive, particularly due to the need for high-resolution images for accurate joint point identification, which limits its implementation in real-time detection tasks, particularly in uncontrolled scenarios [31]. Despite these limitations, recent progress suggests promising potential for future applications of depth sensing in CVGR systems at the edge.

2.3.2. 3D Lasers

Recent investigations have explored 3D and 4D data acquisition to represent the human body's surface during walking or other movements. This data can be obtained using various technologies, including 3D lasers, radar signals [5], Light Detection and Ranging (LiDAR) technologies [32], among other depth sensors. These approaches aim to obtain point clouds, which are sets of data points in 3D space, each with X, Y, and Z coordinates, that provide fine-grained geometric features of a person's body. Furthermore, 4D point clouds extend 3D point clouds by incorporating time as the fourth dimension, allowing for the capture of both spatial and temporal aspects of the person's gait [33]. Consequently, this enables the capture of dynamic walking patterns over a period, facilitating a more detailed analysis of temporal variations in movement.

Among the aforementioned sensors, LiDAR has emerged as a popular choice due to its reliability in capturing gait information even in challenging situations like low light or occlusions. Recent LiDAR-based models have demonstrated compelling results compared to 2D imaging. One example is LidarGait, proposed by Shen et al. [32], which converts sparse point clouds from LiDAR data into depth maps, enabling the capture of the 3D shape of a person's gait. To address the scarcity of LiDAR gait data, the authors created SUSTech1K, a dataset comprising 25,239 gait sequences from 1,050 subjects, covering various factors like visibility, viewing angles, occlusions, clothing, carrying conditions, and scene variations. Challenges for 3D laser-based gait recognition include occlusions, variations in clothing, changes in viewpoint, and the high cost of data acquisition devices, all of which represent important areas for future research [8].

2.3.3. Stereo Vision

Stereo vision enables the reconstruction of 3D scenes through triangulation. Specifically, it uses two or more cameras to capture images of the same scene from slightly different viewpoints, allowing for the perception of depth maps. In the medical field, researchers have applied stereo vision to gait analysis [34] to automate the analysis of screening tests—namely, the Timed Up and Go (TUG) and Performance Oriented Mobility Assessment (POMA) tests—administered to elderly individuals. Their aim was to fuse 2D data with depth information to identify abnormalities that may indicate health risks or mental health decline, achieving compelling results in real daycare facilities. For gait recognition as a biometric tool, authors in [35] proposed a stereo-vision pipeline that extracts 3D contours from a

person’s silhouette in a video sequence, calculates a “stereo gait feature” (SGF) from these contours, reduces dimensionality using PCA, and then classifies gait patterns using neural networks.

While powerful for 3D perception, stereo vision is not ideal for CVGR systems. This is mainly because it demands significant computational resources for real-time 3D scene reconstruction, a major drawback for systems with limited processing power, such as those deployed in real-world settings [36]. Furthermore, accurate calibration of the stereo camera setup is also critical. Without it, depth information becomes unreliable, leading to significant errors in depth estimation and impacting gait recognition accuracy. In contrast, surveillance cameras (RGB and grayscale) are increasingly common due to their low cost and easy installation. They are now prevalent in urban areas, businesses, and public spaces, with growing use in rural areas, private homes, and even natural environments [37]. Therefore, the present article focuses on 2D imaging-based data acquisition.

3. Feature Representations for Computer Vision-Based Gait Recognition

Extracting features from 2D imaging data is crucial in CVGR systems. This step creates feature representations that capture the unique qualities of a person’s walk. As shown in Figure 3, feature extraction methods fall into two main categories: deep learning (DL) based and handcrafted. DL-based methods automatically extract features from raw gait data (e.g., images, video frames, or silhouettes) using DL models, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or long short-term memory networks (LSTMs). These models can be trained end-to-end, integrating feature extraction and gait classification into a single optimised process. Furthermore, DL-based approaches require substantial computational resources and large amounts of labeled training data. DL models are also used to process and classify handcrafted features, leading to improved gait recognition. Section 5.2 provides a more detailed exploration of DL-based feature extraction and common DL models, including their strengths and weaknesses for gait recognition.

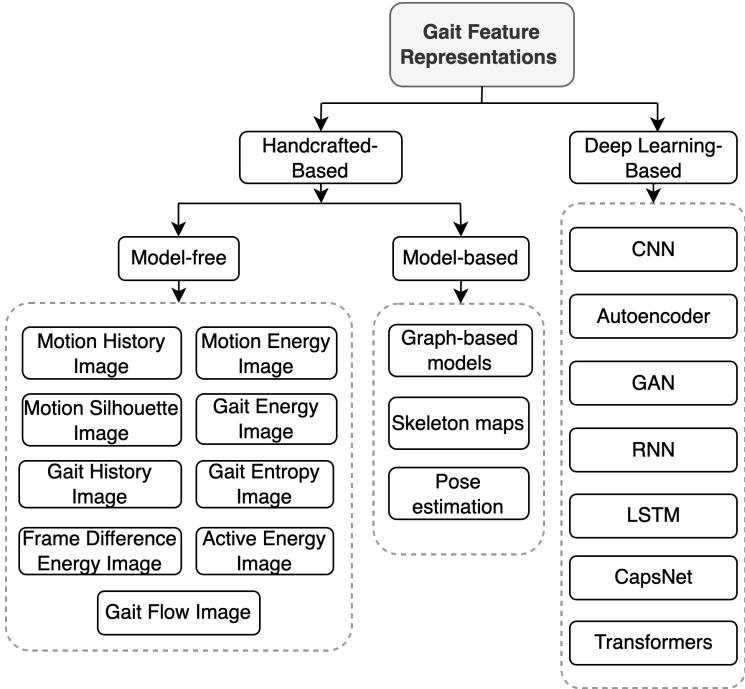


Figure 3. Broad perspective of gait feature representations.

In contrast to deep learning, handcrafted features are manually designed based on researchers’ domain expertise and understanding of distinctive gait characteristics. These features aim to capture the unique structure, motion, and dynamics of an individual’s walking pattern. Handcrafted features can be categorised into two groups: model-based and model-free. Model-based approaches focus on extracting skeletal information and joint positions to capture the localised movement patterns

of each body part [8]. Conversely, model-free methods (also known as appearance-based methods), on the other hand, directly extract features (primarily silhouettes) from a sequence of gait cycles [1]. Table 2 compares handcrafted-based and deep learning-based feature representations. Additionally, Figure 4 illustrates a common CVGR system, including two potential modalities for feature extraction: model-based and model-free. Based on these distinctions, the following sections will delve deeper into model-based and model-free methods.

Table 2. Key differences between handcrafted and deep learning gait feature representations.

Aspect	Handcrafted Features	Deep Learning Features
Feature Design	Manually designed using domain knowledge	Automatically learned from data
Computational Complexity	Generally lower complexity, faster to compute.	Higher computational complexity, requires more resources.
Generalisation Ability	Limited generalisation, often tailored to specific conditions.	Better generalisation, more robust to diverse conditions.
Adaptability	Requires redesign for new tasks or environments.	Learns features adaptively from data.
Data Requirement	Can work with smaller datasets.	Requires large amounts of labeled data for training.
Interpretability	Easier to interpret due to human design.	Harder to interpret, features are learned in black-box fashion.

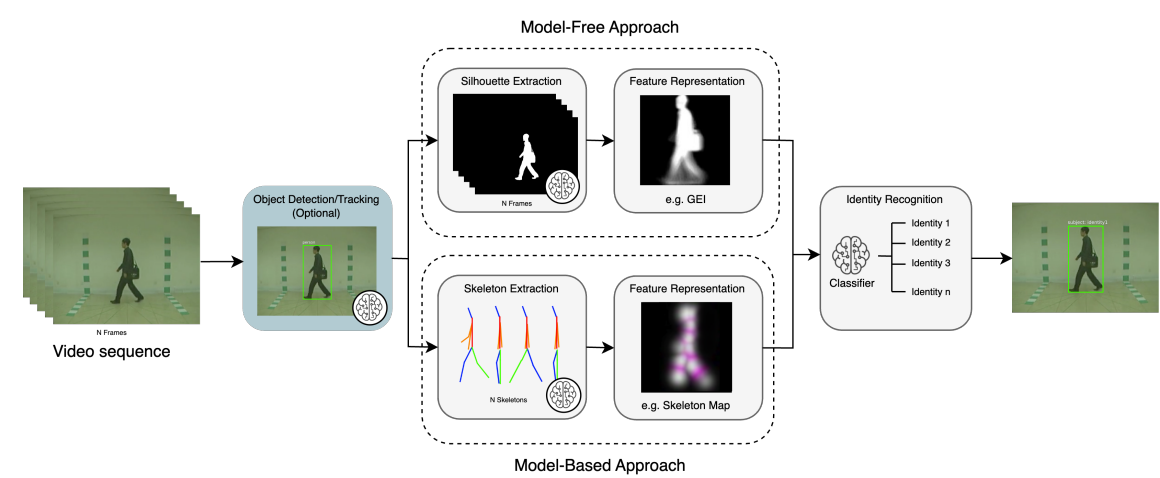


Figure 4. CVGR systems based on handcrafted representations typically employ one of two approaches: the systems extract silhouettes from 2D images (model-free), or rely on human body models (model-based). The video sample shown in the figure comes from the CASIA-B dataset [23].

3.1. Model-Based Representations

Model-based gait recognition requires constructing a mathematical model that represents the human body and its movements during walking. This model can range from a simple stick figure to a complex biomechanical model incorporating muscle activations and joint constraints [8,9]. Other model-based approaches utilise specialised sensors (e.g., Kinect [26] and others mentioned in Section 2.3), or pre-trained models (e.g., OpenPose [38], MediaPipe [39]) to create skeleton-like representations. By extracting spatial and temporal information from human body joints, this approach can distinguish one person’s gait from another. Consequently, they have gained significant attention in recent years for their ability to handle current CVGR challenges such as viewpoint variations, occlusions, and changes in clothing. Table 3 illustrates this rise in investigations of model-based approaches by comparing three important methods in this area.

Table 3. Model-based gait feature representations and their frequency of use in recent research. Identifying the earliest papers introducing each approach can be challenging.

Gait Feature Representation	Year	Pros	Cons	Frequency of Use	Recent Applications
Graph-based Models	2021	Effective for capturing relationships between joints, capable of encoding spatial and temporal dependencies.	Computationally intensive, requires large datasets, can be sensitive to noise in joint detection.	Emerging, increasingly popular	[40–42]
Pose Estimation	2016 (Deep Learning-based)	Captures human joint movements with high granularity, robust to appearance changes, clothing, and background noise.	Sensitive to inaccuracies in joint detection, requires high-quality input, limited in occlusion cases.	Increasingly frequent	[7,31,43]
Skeleton Maps	2010s	Simple representation of body joints, efficient for machine learning models, invariant to appearance changes.	Can miss subtle gait dynamics, reliant on accurate joint detection, struggles in occlusions or missing joint data.	Moderately frequent	[16,43]

3.1.1. Pose Estimation

One prominent model-based technique is pose estimation, which involves detecting human joint positions in 2D or 3D space. This is achieved by using 2D or 3D pose estimation models that detect key joints (e.g., hips, knees, and ankles) in each frame of a video sequence, forming a structured representation of human motion. To improve robustness, recent works employ pose normalisation techniques (e.g., center- and scale-normalisation) that align body joints in a consistent frame of reference, reducing variability caused by camera angles and distances [43]. The following is a sample formulation of pose estimation.

Let the human body pose at time step t be represented by a set of joints $P_t = \{\mathbf{p}_1^t, \mathbf{p}_2^t, \dots, \mathbf{p}_J^t\}$, where $\mathbf{p}_j^t \in \mathbb{R}^2$ or $\mathbb{R}^3$ represents the 2D or 3D coordinates of the j -th joint at time t , and J is the total number of joints. For a sequence of poses across time, the full set of poses is given by $\mathcal{P} = \{P_1, P_2, \dots, P_T\}$, where T is the total number of frames in the sequence. Thus, we can define the joint coordinate representation, where each joint $\mathbf{p}_j^t$ at frame t is represented by its coordinates, depending on whether it is 3D or 2D pose estimation, as shown in Equation (1).

$$\mathbf{p}_j^t = \begin{bmatrix} x_j^t \\ y_j^t \\ z_j^t \end{bmatrix} \quad \text{or} \quad \mathbf{p}_j^t = \begin{bmatrix} x_j^t \\ y_j^t \end{bmatrix} \quad (1)$$

For model-based gait recognition, pose estimation models extract features from the set of joint coordinates. Typically, a feature extraction can be formulated as a function f applied to the sequence of joint positions, as defined in Equation (2), where $\mathbf{F} \in \mathbb{R}^d$ is a d -dimensional feature vector representing the gait. An important step in this methodology is normalisation, which aims to remove variations due to camera angles or distance. This is achieved as shown in Equation (3), where $\mathbf{p}_{\text{center}}^t$ is the centre of the body (e.g., the hip joint), and $\mathbf{p}_{\text{height}}^t$ is the height of the person, used for scaling.

$$\mathbf{F} = f(\mathcal{P}) = f(\{P_1, P_2, \dots, P_T\}) \quad (2)$$

$$\mathbf{p}_j^{t,\text{norm}} = \frac{\mathbf{p}_j^t - \mathbf{p}_{\text{center}}^t}{\|\mathbf{p}_{\text{height}}^t\|} \quad (3)$$

Recent works on pose estimation-based gait recognition have achieved compelling results. For instance, the 3D skeleton-based PoseGait model, proposed by Lia et al. in 2020 [31], uses CNNs to extract multi-dimensional gait features from body joint trajectories. More recently, in 2023, Fu et al. introduced the novel framework GPGait [7] to address pose-based models' main limitation: their inability to generalise across different scenarios. GPGait utilises two types of gait feature representations: Human-Oriented Transformation (HOT) and Human-Oriented Descriptors (HOD). HOT aligns human pose data from various camera viewpoints into a unified, stable representation, while HOD comprises multi-branch feature extraction modules that generate discriminative, domain-invariant features from this unified pose. Finally, to classify features, the authors employed a Part-Aware Graph Convolutional Network (PAGCN), which efficiently implements graph partition and constructs local-global relationships through mask operations on the adjacency matrix.

3.1.2. Skeleton Maps

Recent progress in skeleton maps for model-based gait recognition has focused on improving the precision and compactness of gait representations derived from human joint coordinates. For instance, SkeletonGait [43] generates heatmaps from joint coordinates, aiming to eliminate gait-unrelated information (e.g., walking trajectory, filming angle) through data normalisation techniques. This includes centre- and scale-normalisation to filter out irrelevant information like walking trajectory and camera distance, ensuring a cleaner representation of gait features. Moreover, newer methods in this

field further refine skeleton maps by leveraging only essential gait-related details, improving accuracy on challenging outdoor datasets like OUMVL [44] and GREW [16].

The representation of skeleton maps can be formulated as follows: Let the human skeleton at frame t be represented as a skeleton map S_t . This map consists of a set of joint positions $S_t = \{\mathbf{s}_1^t, \mathbf{s}_2^t, \dots, \mathbf{s}_J^t\}$, where $\mathbf{s}_j^t \in \mathbb{R}^2$ or $\mathbb{R}^3$ denotes the position of the j -th joint at frame t . The joint feature vector can then be defined as a collection $\mathbf{s}_j^t$ augmented with additional features such as velocity, v_j^t , and angles, θ_j^t , at joint j at time t , as denoted in Equation (4).

$$\mathbf{h}_j^t = \begin{bmatrix} x_j^t \\ y_j^t \\ z_j^t \\ v_j^t \\ \theta_j^t \end{bmatrix} \quad (4)$$

Using the joint feature vector, we can construct the feature matrix for the skeleton map at frame t , as shown in Equation (5), where d is the dimension of the joint feature vector. This matrix effectively captures the spatial and temporal characteristics of the human pose at each frame. (Note that Figure 4 illustrates a common CVGR system that utilises these features.)

$$\mathbf{H}^t = \begin{bmatrix} \mathbf{h}_1^t \\ \mathbf{h}_2^t \\ \vdots \\ \mathbf{h}_J^t \end{bmatrix} \in \mathbb{R}^{J \times d} \quad (5)$$

3.1.3. Graph-Based Models

Graph-based models for model-based gait recognition have seen significant advancements in recent years, primarily leveraging Graph Convolutional Networks (GCNs) to better capture spatial and temporal dependencies in human motion. For instance, in 2023, GaitGraph [40] and its successor, GaitGraph2 [41], treat human joints as nodes and their connections (limbs) as edges, processing this data to capture both local and global spatial relationships. This approach enables the capture of both local movements (e.g., individual limb movements) and global motion patterns (e.g., full-body gait dynamics), crucial for distinguishing gaits across individuals. These models have demonstrated improved robustness across various datasets and under different environmental conditions, particularly in handling occlusion, different viewing angles, and variations in walking surfaces. Moreover, their lightweight nature makes them well-suited for real-time gait recognition, even in resource-constrained environments such as mobile devices or edge computing platforms.

The mathematical representation of a graph-based model for gait recognition can be formulated as follows. A human skeleton at frame t can be represented as a graph $G_t = (V, E)$, where $V = \{v_1, v_2, \dots, v_J\}$ is the set of J joints (nodes), and $E \subseteq V \times V$ is the set of edges representing the physical connections between joints (limbs). Each node v_j at frame t is associated with a feature vector $\mathbf{h}_j^t \in \mathbb{R}^d$, representing the position (e.g., 2D or 3D coordinates) and possibly other information such as velocity or joint angles:

$$\mathbf{h}_j^t = \begin{bmatrix} x_j^t \\ y_j^t \\ z_j^t \end{bmatrix} \quad (\text{for 3D pose}). \quad (6)$$

Thus, the feature matrix for the entire skeleton at frame t is:

$$\mathbf{H}^t = \begin{bmatrix} \mathbf{h}_1^t \\ \mathbf{h}_2^t \\ \vdots \\ \mathbf{h}_j^t \end{bmatrix} \in \mathbb{R}^{J \times d} \quad (7)$$

The connections between joints (limbs) are captured by the adjacency matrix $\mathbf{A} \in \mathbb{R}^{J \times J}$, where:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{if there is an edge between joints } v_i \text{ and } v_j, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

3.2. Model-Free-Based Representations

Model-free representation methods for gait recognition, also known as appearance-based gait recognition, extract features directly from gait data, without relying on explicit models of the human body or specific anatomical configurations. This direct extraction makes them practical for 2D gait data and particularly well-suited for edge computing due to their efficiency in online processing pipelines. However, it is important to highlight the decreasing trend in research on these types of representations due to the rise of model-based methods and deep learning.

3.2.1. Motion Energy Images and Motion History Images

Human motion analysis has long been an area of interest for computer vision researchers. Bobick and Davis [45] proposed two of the initial feature representations: the Motion-Energy Image (MEI) and the Motion-History Image (MHI). An MEI is a binary image that highlights where motion has occurred within a certain time window, with pixels taking a value of 1 to indicate motion, and 0 to indicate no motion. Equation (9) defines this, where $D(x, y, t)$ represents a binary image sequence indicating regions of motion, with x and y as the spatial coordinates for frame t , and τ as the temporal extent of the motion.

$$E_\tau(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i) \quad (9)$$

Conversely, the MHI aims to produce a scalar-valued image where more recently moving pixels appear brighter. Equation (10) defines MHI, where H_{tau} is a function representing the temporal history of motion at a specific pixel. Inspired by MHI, Lam and Lee proposed a derived gait representation called the Motion Silhouettes Image (MSI) [46]. In their proposal, they replaced τ with 255, and made MSI simpler and easier to implement, maintaining most of functionality of the MHI [10] under similar conditions.

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t - 1) - 1) & \text{otherwise} \end{cases} \quad (10)$$

3.2.2. Gait Energy Images and Gait History Images

Gait Energy Images (GEI) are 2D representations of a person's gait over time, capturing their movement into a single composite image [47]. They are often created by averaging or accumulating a series of aligned silhouettes, motion images (such as MSIs), or depth images. As shown in 4, a GEI provides a holistic view of an individual's walking style, capturing key characteristics like stride length, arm swing, and body posture within a single image, while preserving all temporal information. Equation (11) shows the formulation of a GEI, where $B_t(x, y)$ represents a binary gait silhouette image at frame t , and n is the total number of frames within the cycle. While the GEI captures both dynamic

(movement-related) and static (shape-related) features from the gait sequence, it lacks the ability to characterise the specific timing of foreground activity at each pixel. To address this, an improved temporal feature representation method called the Gait History Image (GHI) was proposed by Liu and Zheng [48]. The GHI is designed to represent not only static and dynamic characteristics but also the spatial and temporal variations within a gait cycle. Equation (12) defines this, where $S(x, y)$ represents static pixels, calculated by taking the intersection of all motion detection images $D(x, y, t)$ from time 1 to τ .

$$G(x, y) = \frac{1}{n} \sum_{t=1}^n B_t(x, y) \quad (11)$$

$$E_{GHI}(x, y) = \begin{cases} \tau & \text{if } S(x, y) = 1 \\ \sum_{t=1}^t D(x, y, t) \cdot (t - 1) & \text{otherwise} \end{cases} \quad (12)$$

Both GHI and GEI are widely used methods in modern gait recognition workflows. However, their improvement remains an active area of research. For example, Zebhi et al. [49] proposed combining GHI with their new descriptors named Time-sliced Averaged Gradient Boundary Magnitude (TAGBM) to abstract spatial and temporal information from videos into templates for human activity recognition. Similarly, Wang et al. [50] recently presented DyGait, a CVGR system that improves traditional GEIs. They introduced a Dynamic Augmentation Module (DAM) that enhances the traditional GEI approach by focusing on the dynamic aspects of human gait. While traditional GEI primarily captures static body parts, like the torso, the DyGait method augments this by emphasising the dynamic movements of limbs, allowing for more robust gait recognition, especially in challenging conditions like changes in clothing or when individuals are carrying items.

3.2.3. Gait Entropy Images

Unlike GEIs, Gait Entropy Images (GENI) focus on the dynamic elements of gait by encoding the randomness of pixel values within the silhouette over a complete gait cycle [51]. GENI are constructed by calculating the Shannon entropy (defined in Equation (13)) for each pixel over a complete gait cycle. This means that pixels with high values in the GENI correspond to body parts with high movement variability, such as the arms and legs.

$$H(x, y) = - \sum_{k=1}^{k_n} p_k(x, y) \log_2 p_k(x, y) \quad (13)$$

In the previous equation, $p_k(x, y)$ is the probability that the pixel at the coordinates x, y takes on the k^{th} value. For binary images, there are only two possible intensity values, so $k_n = 2$. Therefore, the gait entropy image $G(x, y)$ can be calculated in function of a scaled and discretised $H(x, y)$, obtained as follows:

$$G(x, y) = \frac{(H(x, y) - H_{min}) \times 255}{H_{max} - H_{min}} \quad (14)$$

where H_{min} represents the minimum value of $H(x, y)$, and H_{max} represents the maximum value. The resulting frame will highlight the dynamic areas of the human body with higher intensity values, while static parts will have lower intensity values.

3.2.4. Gait Flow Image

Gait Flow Image (GFI) aims to capture the dynamic motion information in a gait sequence by utilizing optical flow [52]. Optical flow algorithms estimate the apparent motion of pixels between consecutive frames in a video sequence. For gait movement patterns, this provides a compelling means to capture the subtle dynamics of human movement, such as the swinging of limbs, the shifting of body weight, and the overall rhythm of walking. The GFI is constructed by first extracting silhouettes of the

walking person from a sequence of images and then calculating the optical flow between consecutive frames in the silhouette sequence. The primary formula involved in calculating optical flow is the Optical Flow Constraint, defined as follows:

$$I_x u + I_y v + I_t = 0 \quad (15)$$

In this equation, I_x represents the spatial derivative of the image intensity in the x-direction (horizontal), I_y represents the spatial derivative of the image intensity in the y-direction (vertical), I_t represents temporal derivative of the image intensity (change in intensity over time). Furthermore, u and v represent the horizontal and vertical component of the optical flow vector, respectively.

Optical flow qualities were recently explored by Ye et al. [53], who implemented Gait Optical Flow Image (GOFI) - a subtype of GFI where the optical flow information is directly encoded into the image from RGB frames - to add the instantaneous motion direction and intensity to original gait silhouettes. They then implemented a neural network called the Gait Optical Flow Network (GOFN), which allows for a more detailed and robust analysis of gait patterns. Despite its qualities, optical flow requires high-computational power to process resolution videos or for real-time applications, which can limit the scalability and efficiency of gait recognition systems, particularly on resource-constrained devices [8].

3.2.5. Other Model-Free Representation Methods

This section summarises other noteworthy mode-free gait representations. Additionally, Table 4 compares all the feature representations introduced in this section.

- To address the issue of poor image segmentation in gait recognition frameworks, Chen et al. [54] proposed the Frame Difference Energy Image (FDEI) representation. This is a robust representation designed to mitigate the impact of incomplete silhouettes.
- Wang et al. [55] proposed the Chrono-Gait Image (CGI), which encodes the temporal information by assigning different colours to the silhouettes based on their position in the gait cycle, generating a single CGI with richer information.
- To tackle the issue of variations in clothing and carried objects during walking, Zhang et al. [56] proposed the Active Energy Image (AEI) representation. This approach focuses on the dynamic body parts, discarding the static ones, by calculating the difference between consecutive silhouettes in a gait sequence.
- He et al. [57] proposed the Period Energy Image (PEI), a multichannel gait template designed as a generalisation of GEI. It aims to enrich spatial and temporal information in cross-view gait recognition, maintaining more of this information compared to other templates.
- The frame-by-frame Gait Energy Image (ff-GEI) presented in [58] effectively expresses available gait data, relaxes the gait cycle segmentation constraints imposed by existing algorithms, and is better suited to the requirements of DL models.

Table 4. Comparison of model-free gait feature representations, sorted by year of publication.

Gait Feature Representation	Year	Pros	Cons	Frequency of Use	Recent Applications
ff-GEI [58]	2020	Captures energy per frame, useful for detailed gait dynamics.	Computationally expensive, large data requirements	Rare	No recent works found.
PEI [57]	2019	Highlights periodic gait motion, useful for recognising consistent gait patterns.	Sensitive to changes in walking speed and conditions.	Less frequent	[59]
GFI [52]	2011	Captures velocity and flow of movement, sensitive to gait dynamics.	Computationally complex, sensitive to noise and illumination changes.	Less frequent	[53], [8]
CGI [55]	2010	Combines motion with temporal encoding, captures both spatial and temporal information.	Computationally complex, sensitive to frame rate and noise.	Less frequent	No recent works found.
AEI [56]	2010	Captures active energy regions, good for detecting dynamic motion.	Complex to compute, requires high-quality input for effectiveness.	Less frequent	[60]
FDEI [54]	2009	Highlights regions of change between frames, simple to compute.	Misses subtle motions, highly sensitive to noise	Moderately frequent.	[61]
GEnI [51]	2009	Encodes gait variability and randomness, useful for capturing subtle dynamics.	Sensitive to noise, more complex to compute.	Less frequent	[62]
GHI [48]	2007	Captures both spatial and temporal aspects of movement.	More computationally expensive, sensitive to noise	Less frequent.	[49]
GEI [47]	2006	Robust to clothing and carrying conditions, captures averaged body silhouettes.	Loses fine temporal details, less effective in occlusion scenarios.	Very frequent	[50]
MSI [46]	2005	Simpler and easier to implement than MHI.	Primarily focuses on shape information without incorporating the temporal dynamics of the gait.	Moderately frequent	No recent works found.
MEI [45]	2001	Simple and efficient, captures where motion has occurred.	Lacks detailed temporal motion information, sensitive. to noise	Moderately frequent	No recent works found.
MHI [45]	2001	Captures temporal motion patterns, simple to compute, compact representation.	Sensitive to noise, cannot capture subtle variations in movement.	Less frequent	No recent works found.

4. Gait Representation Dimensionality Reduction

Dimensionality reduction (DR) techniques play a vital role in gait recognition by eliminating redundant or irrelevant features from gait representations. This minimisation process offers several benefits: it reduces computational costs, filters out noise, and improves the accuracy of subject recognition [3]. While DL encoders are now commonly used for DR nowadays, the resulting feature representations for gait sequences can still be quite high-dimensional. To address this, DR methods can reduce the number of features, making subsequent classification or recognition tasks more manageable [63]. This is particularly important because processing high-dimensional features can be computationally demanding, especially for real-time applications or resource-constrained devices. Therefore, DR helps alleviate this burden, enabling efficient edge computing.

4.1. Linear Techniques

One of the most commonly used DR techniques in gait recognition is Principal Component Analysis (PCA) [64,65]. PCA aims to identify the optimal projection that maximises the variance of the information, effectively reducing the dimensionality of feature vectors. This technique compresses large data representations into smaller feature sets that still retain the crucial information of the original representation [66]. For instance, in the framework proposed by Gupta et al. [67], a Boundary Energy Image (BEI) served as the feature representation. PCA was then applied for DR before using a Linear Discriminant Analysis (LDA) algorithm for gait classification.

Regarding LDA, its main goal is to find a new set of axes (linear combinations of the original features) that maximise the separation between different classes while minimising the variance within each class. This transformed space, with fewer dimensions, makes it easier to classify new data points. LDA has been applied in numerous CVGR frameworks. For example, Guo et al. [68] implemented LDA to the feature data reduction alongside a Gabor Filter for feature extraction from the GEI and used an Extreme Learning Machine algorithm for recognition and classification. Additionally, Wang et al. [69] introduced a generalised LDA based on the Euclidean norm (ELDA) to address the challenge of indistinguishability resulting from overlapping data in LDA, and employed multi-class SVMs for gait classification. Furthermore, in [70], the authors presented a DR framework based on LDA and PCA in conjunction with a GEI representation and a Cyclic Recognition Code (CRC) classification scheme for video-based gait recognition.

4.2. Non-linear Techniques

Non-linear techniques for gait representation DR, such as Isomap [71], Locally Linear Embedding (LLE) [72], and Laplacian Eigenmaps [73], operate under the assumption that gait data exists on a low-dimensional manifold embedded in a high-dimensional space. These techniques aim to preserve the local and global structure of gait data while reducing its dimensionality. Another valuable technique, t-distributed Stochastic Neighbor Embedding (t-SNE), is commonly used for visualising high-dimensional data. By focusing on preserving local relationships between data points, it is useful for exploratory data analysis and identifying clusters in gait recognition tasks. For example, Che and Kong [74] investigated how t-SNE can be applied to GEI images of walking subjects after the images underwent Discrete Wavelet Decomposition (DWT) to reduce their dimensionality. Their results demonstrated that this proposed feature extraction method effectively reduces computational complexity and preserves image information, leading to improved precision metrics when using an SVM model.

Although non-linear techniques have shown promise in gait recognition, they are less frequently used in state-of-the-art DR for gait representation compared to linear techniques. This is primarily because most non-linear techniques can be computationally more expensive than linear methods, especially when dealing with large datasets. Moreover, this high computational cost can significantly limit their implementation on embedded devices. However, manifold learning techniques may

still offer advantages when data exhibits strong non-linear relationships or when interpretability is crucial [75].

4.3. Other Dimensionality Reduction Methods for Gait Recognition

The Discrete Cosine Transform (DCT) is a widely used technique for DR in image-based gait recognition. It retains crucial gait features, such as body posture and limb movement, by utilising a limited set of low-frequency DCT coefficients. Prior research has demonstrated the effectiveness of DCT in reducing image dimensionality before classification tasks, where a significant number of DCT coefficients can be discarded, substantially decreasing the search space dimensionality [76]. In the context of gait recognition, Fan et al. [77] proposed an approach for DR that combines DCT and LDA. This method effectively leverages the frequency-domain information of the GEI representation to achieve higher recognition rates than other reviewed methods. In another study by Chhatrala et al. [78], the authors have combined the Gabor function and DCT to extract features from gait representations, resulting in the extraction of highly discriminative components that led to high recognition accuracy.

The combination of two or more DR methods for gait recognition is prevalent in the literature. Similarly to Che et al. [74], who combined t-SNE with DWT, Shivani and Singh [79] proposed a method for reducing redundant data and extracting features by combining DWT and DCT prior to classification. This approach effectively extracts critical features from gait images and video frames. Furthermore, Wen and Wang [80] introduced a gait recognition technique using sparse linear subspace to reduce the dimensionality of frame-by-frame gait energy images (ff-GEIs) generated from primary gait features. This approach was tested on two well-known gait datasets and demonstrated its effectiveness.

5. Classification of Gait Feature Representations

The final stage in gait recognition involves classifying the previously extracted and reduced gait feature representations. This section reviews the various classifiers commonly employed in this process.

5.1. Traditional Models

5.1.1. Distance-Based Classification

Euclidean distance (Equation (16)) is a simple and widely used method for classifying gait features in CVGR systems [81–83]. It compares and matches gait representations by calculating the distance between them. Another popular approach is the k-Nearest Neighbours (k-NN) algorithm, which classifies a new gait sample based on the majority class among its 'k' closest neighbours in the feature space. Neighbouring samples are defined using metrics such as Euclidean distance, Manhattan distance, or cosine similarity (Equations (16), (17), and (18), respectively). Several studies have successfully employed k-NN in gait recognition. For example, Pratama et al. [84] combined a k-NN classifier with GEI, Gabor wavelets, and PCA. Additionally, Premalatha and Chandramani [85] utilised k-NN classification along with GEI partitioning and region selection.

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (16)$$

$$d(p, q) = \sum_{i=1}^n |q_i - p_i| \quad (17)$$

$$\cos(p, q) = \frac{p \cdot q}{\|p\| \|q\|} \quad (18)$$

5.1.2. Machine Learning

Traditional machine learning techniques for CVGR systems encompass a diverse set of models, among which Support Vector Machines (SVMs) have been a particularly popular choice for classification. SVMs, when applied to gait recognition, aim to map gait representations to a high-dimensional feature space where they can be effectively categorised [86]. Several notable contributions using SVMs for CVGR can be highlighted. For instance, Wattanapanich et al. [64] proposed a view-invariant gait recognition framework that utilised a One-Against-All SVM classifier. This framework sought to achieve robustness and address the challenges posed by unknown camera view angles and appearance changes. Their results demonstrated the framework's robustness under varying camera view angles. In another study, Wang et al. [87] presented a gait recognition algorithm employing a multi-class SVM. This algorithm leveraged Gabor wavelets for gait feature extraction and Two-directional Two-dimensional PCA ((2D)² PCA) for dimensionality reduction.

Hidden Markov Models (HMMs) are prominent in gait analysis due to their ability to model the sequential and temporal dynamics of gait data, capturing variations in the gait cycle over time [88]. These traditional methods, including LDA [67,89], Bayesian Classifiers [90,91], often perform well with model-free features and have laid the groundwork for the deep learning techniques that now dominate the field. However, model-based features often involve more complex data, such as joint angles, bone lengths, or motion parameters. This can lead to higher dimensionality, potentially making it harder for traditional models discriminate effectively [92].

5.2. Deep Learning

Deep Learning (DL), a subfield of machine learning, utilises artificial neural networks (ANNs) with multiple layers to learn complex patterns from data. These deep neural networks (DNNs), inspired by the human brain, have proven effective in tackling gait recognition challenges such as intra-class variations (e.g., carrying objects, wearing different clothes) [93] and cross-view recognition [94]. Consequently, CVGR pipelines based on DL can be clustered into two groups: end-to-end and modular. In end-to-end systems, the encoder component of DL models enables CVGR systems to extract subtle gait features, while the DNN deciphers complex patterns that were previously elusive. Figure 5 illustrates this, showing that a DL model can be fed with raw gait data, frames containing silhouettes or skeletons modelling the human body of the walking subject. Although silhouettes or skeletons can be obtained by using specific algorithms or DL models, the end-to-end approach focuses on training a single neural network to handle the entire gait recognition task, from raw input to final output.

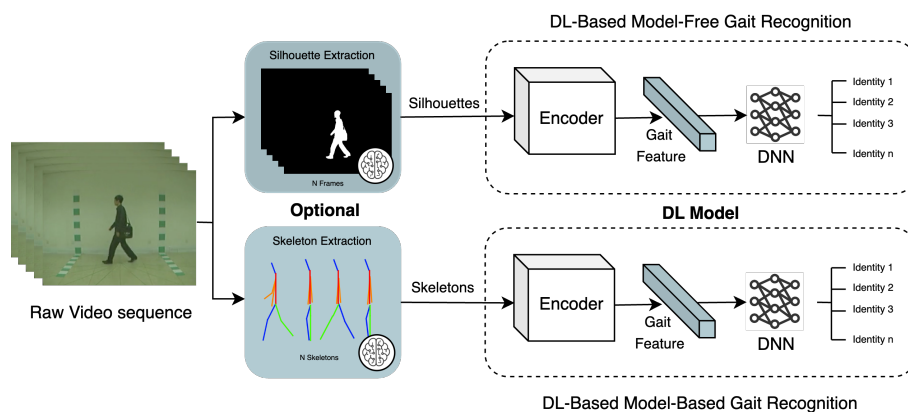


Figure 5. Two types of end-to-end gait recognition pipelines based on model-based and model-free feature representations.

In contrast, modular pipelines might also include handcrafted feature extraction and dimensionality reduction modules (detailed in Section 3 and 4, respectively), object detectors (to detect people), semantic segmentation models (to extract silhouettes), and image classifiers (to recognise

identities) [95], making the CVGR system more complex to deploy on a single low-resource device. Figure 4 shows how these tasks fit into the typical gait recognition pipeline, with brain symbols marking tasks suitable to be performed by individual DL models. Furthermore, both end-to-end and modular pipelines can incorporate object tracking to follow multiple subjects. The following sections explore representative DL models used frequently in CVGR systems.

5.2.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a specialised type of DNN designed primarily for processing and analysing visual imagery. Inspired by the visual cortex in humans, they excel at image recognition. Their key distinguishing feature is their ability to recognise patterns regardless of where they appear in the image [96]. This makes them particularly useful for model-free approaches, where convolutional layers extract features from gait representations, and subsequent hidden layers learn to identify the most salient information. As Figure 1b shows, CNNs are one of the most frequently used models for gait recognition. For example:

- In 2019, Min et al. [97] presented a simple ten-layer CNN architecture that processes GEI representations as input. Their architecture consists of convolutional, pooling, and fully connected layers. The authors compared different activation functions for training a multi-view angle gait recognition model and found that a CNN with LeakyReLU slightly outperformed others in terms of average accuracy and correct classification rates (CCRs).
- In 2022, Ambika et al. [98] proposed a CNN-MLP (Multilayer Perceptron) based approach for gait classification, aiming to be robust to velocity variations and appearance covariates such as carrying a backpack.
- Further CNN-based gait recognition approaches were proposed in [59] and [99].

5.2.2. Autoencoders

An autoencoder is a type of ANN that consists of an encoder and a decoder, with the primary goal of reconstructing its input as accurately as possible using the decoder module [100]. Both encoder and decoder modules can have convolutional or deconvolutional layers to better process visual imagery, and the bottleneck layer (middle layer) typically serves as a dimensionality-reduced representation. Autoencoders have been utilised in model-free pipelines to address appearance change issues like varying view angles [101] and for completing missing parts in GEI representations [102]. On the other hand, recent compelling investigations for model-based methods include:

- Guo et al. [103], who recently introduced a physics-augmented autoencoder (PAA) that integrates a physics-based decoder into the architecture. By incorporating physics, the learned 3D skeleton representations become more compact, and physically plausible.
- Li et al. [104], who recently proposed a novel gait recognition method that leverages the Koopman operator theory and invertible autoencoders to improve interpretability and reduce the computational cost of gait recognition by learning a low-dimensional, physically meaningful representation of gait dynamics that captures the complex kinematic features of gait cycles.

5.2.3. Generative Adversarial Networks

Similarly to autoencoders, Generative Adversarial Networks (GANs) are a powerful DL approach capable of producing new data that closely resembles the training data. In the context of gait recognition, various GAN approaches have been adopted to address view variations or appearance changes. For example:

- Yu et al. [105] proposed a GAN to tackle gait recognition limitations related to appearance changes by generating realistic-looking gait images. The results demonstrated excellent performance on large datasets, where a canonical side-view gait image was generated without needing prior knowledge of the subject's view angle, clothing, or carrying conditions.

- To further research the challenge of view variations in gait recognition, Zhang et al. [106] proposed the View Transformation GAN (VT-GAN). This model translates gaits between any two views using an identity preserver module to prevent the loss of personal identity information during transformations.

5.2.4. Capsule Neural Networks

Capsule networks, or CapsNets, are a type of DNN designed to address some limitations of traditional CNNs, particularly in understanding hierarchical relationships and spatial information within images [107]. Capsules, the fundamental building block of CapsNets, are groups of neurons that encapsulate information about specific features or entities in the input data. Specifically, each capsule in a CapsNet comprises a set of neurons whose outputs represent different attributes of a single feature. CapsNets have been successfully used in gait recognition to mitigate performance drops when dealing with various walking conditions or multiple viewing angles. For example, Sepas-Moghaddam et al. [108] aimed to address the gait recognition covariate problem by proposing a DNN for learning to transfer multiscale partial gait representations. This network employs CapsNets to learn deeper part-whole relationships and assign more weight to highly relevant features, while disregarding spurious dimensions. The proposed model consists of a partial feature extraction stage for gait maps, a BGRU layer, a capsule attention module, and a final layer with a softmax activation function for classification. The model demonstrated impressive rank-1 recognition results, achieving mean accuracy values of 95.7% for normal walking, 90.7% for carrying a bag, and 72.4% for varying clothing conditions. Other works, such as [109–111] have also incorporated capsule modules into their gait recognition approaches.

5.2.5. Recurrent Neural Networks

Recurrent Neural Networks (RNNs), an extension of DNNs well-suited for processing sequential data, offer another approach to gait recognition classification [112]. These models, along with Long Short-Term Memory (LSTM) models – a type of RNN adept at learning order dependency in sequence prediction – have been widely used for gait recognition by leveraging their inherent ability to capture temporal dependencies. Although RNNs and LSTMs seem like a natural fit for gait recognition due to their ability to handle sequences, they have largely been overshadowed by CNNs, primarily because of their computational cost and limitations in capturing spatial details [8,113]. This is particularly relevant for resource-constrained environments or applications demanding real-time performance, where RNNs/LSTMs might not be the most practical choice [114,115]. Nevertheless, they can still be valuable when capturing temporal dynamics is paramount, especially when combined with CNNs to leverage the strengths of both. Some remarkable works in this field include:

- Wang and Yan [58], who combined LSTM with convolutional layers (namely ConvLSTM) and trained it on ff-GEI gait feature representations. Their method outperformed several state-of-the-art works in multi-view angle gait recognition.
- To address the challenge of occlusions in gait recognition, Sepas-Moghaddam et al. [116] proposed a method for learning invariant convolutional gait energy maps with an attention-based recurrent model. Their network structure utilised partial representations by decomposing learned gait representations into convolutional energy maps. A recurrent learning module composed of bidirectional gated recurrent units (BGRU) was then used to exploit relationships between these partial spatiotemporal representations, coupled with an attention mechanism to focus on crucial information.

5.2.6. Graph Neural Networks

Graph Neural Networks (GNNs) are a class of DL models designed to process data that can be represented as graphs. As explained in Section 3.1.3, gait features (e.g., skeleton maps) can also be represented as graphs, making it possible to combine them with GNNs for efficient gait recognition.

Unlike traditional ANNs, which typically operate on structured data (like grids or sequences), GNNs can naturally handle non-Euclidean data by aggregating and passing information between neighbouring nodes. This allows GNNs to capture complex relational structures and dependencies within the gait feature data [40]. Crucially, GNNs have been shown to be capable of running on edge devices with appropriate optimisation techniques listed in Section 7 [117].

Graph Convolutional Networks, a subtype of GNNs, are designed to process graph data by applying convolutions. In gait recognition, GCNs can capture the complex relationships between joints in a gait sequence. Both GNNs and GCNs have recently shown great promise by modelling the structural and temporal dependencies of human body joints and keypoints for gait recognition. For example, Guo et al. [118] proposed a notable work in this field that uses a Spatial-Temporal Graph Convolutional Network (ST-GCN) to capture both spatial joint relationships and temporal walking dynamics. These and other aforementioned GNN-based approaches, such as [7,41], have successfully leveraged the inherent structural and dynamic properties of human gait, leading to state-of-the-art results on benchmark datasets like CASIA-B [23] and OU-MVLP [44].

Recently, GNNs have also obtained meaningful results with 3D and 4D point clouds. For example, Ma et al. [119], proposed the Dynamic Aggregation Network (DANet), a network based on the Dynamic Graph CNN (DGCNN) [120]. DANet operates on dynamic graph structures and uses edge convolution to learn both the local and global structure of the point cloud, which is crucial for gait recognition from dynamic 3D data. Specifically, DANet dynamically learns local and global motion patterns from gait sequences, enhancing the model's ability to focus on key regions within the gait data and achieving compelling results on datasets such as Gait3D [5].

5.2.7. Transformers and Attention Mechanisms

Transformer networks excel at capturing long-range dependencies in sequential data and are effective at processing gait in its different forms, where temporal variations are key [121]. Furthermore, attention mechanisms enable transformers to focus on specific parts of complex inputs, thereby increasing their performance. Some compelling works in this field include the following:

- To enhance discrimination between different classes of gait features, Wang and Yan [122] (2021) presented a self-attention-based classification model that combines non-local and regionalised features. This combination helps identify relevant non-local features, which are then refined by a two-channel network.
- Jia et al. [123] (2021) proposed a CNN Joint Attention Mechanism (CJAM) for identifying the most crucial pixels in a gait sequence. Their workflow uses a CNN to extract feature vectors from the initial gait frames and feeds them into an attention model comprising encoder and decoder layers, followed by linear operations for information transformation and a softmax function for classification. Twelve experiments demonstrated that this attention model outperforms others in terms of reducing errors, with the CJAM model achieving accuracy improvements of 8.44%, 2.94%, and 1.45% over 3D-CNN, CNN-LSTM, and a simple CNN, respectively.
- Mogan et al. [124] (2022) converted GEI representations into 2D flattened patches and passed them into a Vision Transformer model consisting of an embedding layer (with patch embedding applied to the sequence of patches), a transformer encoder for achieving a final representation, and a multi-layer perceptron that performs the classification based on the first token of the sequence. Experiments on small and large datasets demonstrated the scalability of the Visual Transformer model, showcasing its robustness against noisy and incomplete silhouettes and its remarkable subject identification capability regardless of the different covariates.
- Li et al. [121] (2023) proposed TransGait, a novel gait recognition framework that utilises a set transformer to effectively fuse silhouette and pose information, leveraging the strengths of both feature representations.
- Researchers have also recently leveraged self-attention [124], spatial-temporal attention [125] to improve the discriminative power of gait features.

In summary, the CVGR field is actively exploring novel models to enhance gait recognition performance in challenging real-world scenarios. However, model complexity can hinder deployment on low-resource devices, often necessitating cloud-based premises. For instance, although deploying transformers on embedded devices is possible [126], they can be quite resource-intensive, demanding substantial memory for parameters and intermediate computations. Similarly, the complex computations of many DL models, like transformers, often require significant processing power. Table 5 further analyses this by providing a comparison that describes the suitability of different DL models for edge computing. This highlights the importance of distributed inference edge-computing architectures, which aim to collect data and make detections closer to the data source.

Table 5. Comparison of DL models for gait recognition and suitability for edge computing.

Model	Frequency of Use in CVGR Systems	Suitability for Edge Computing	Recent Applications for Edge Computing
CNNs	High	Moderate, requires optimization	[127,128]
Autoencoders	Moderate	High, compact representations	[129]
GANs	Moderate	Low, resource-intensive	[130]
CapsNets	Low	Low, high complexity	[131]
RNNs	High	Moderate, LSTM/GRU optimizations required	[132]
GNNs	Moderate	Low, requires significant optimization	[117]
GCNs	Moderate	Low, optimization needed for real-time deployment	[133]
Transformers	Increasing	Low, requires optimization for edge devices	[126]

6. Edge-Oriented Inference Architectures

Edge computing is rapidly evolving and gaining traction across industries. Advancements in hardware, software, and AI algorithms are making intelligent devices increasingly powerful, allowing for the deployment of more sophisticated AI applications directly at the edge. This trend, also known as Edge AI, is expected to continue, leading to wider adoption of intelligent systems in the years to come [134]. Within this context, three main strategies for edge computing deployment can be observed in the current literature: Edge server-based, End device-based, and a hybrid approach combining Edge server + End devices or Edge server + Cloud server (refer to Figure 6 for a visual representation). As shown in Figure 7, these approaches are not independent of each other and can be combined simultaneously to enable more feasible real-time computation in CVGR systems. The following sections will describe these strategies in detail.

6.1. Edge Server-Based

Unlike cloud-based servers, edge servers are located at the “edge” of a network, close to where data is collected. They vary in form factor, ranging from powerful rack-mounted servers to high-capability desktop computers. In this model, end devices collect data and transmit it directly to an edge server, which processes the input and send the results back. The inference pipeline on the edge server may include one or more steps for pre-processing, recognising, and post-processing visual data, and the server must be able to handle visual data from multiple end devices simultaneously. However, a primary drawback of this model is its reliance on network bandwidth; processing cannot occur without a local area network. Galanopoulos et al. [135] provide an example of this approach, where mobile devices were programmed to collect video data and send it to an edge server for real-time processing and analytics.

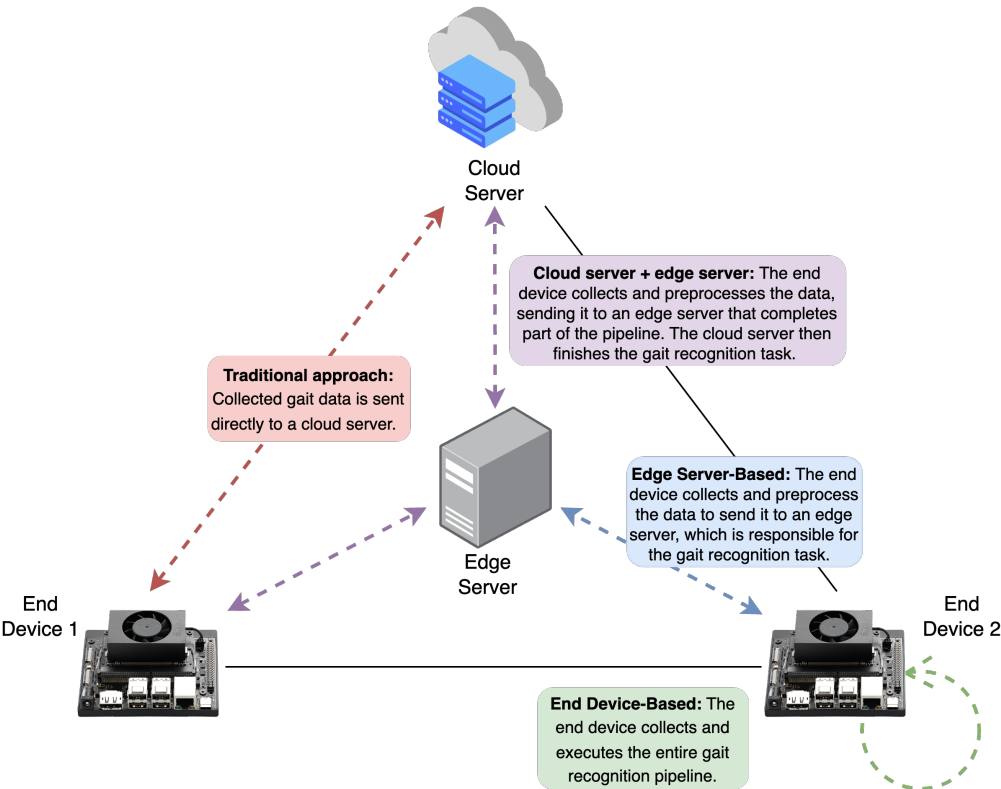


Figure 6. Graphical depictions of various edge-oriented inference architectures, considering the data lifecycle from capture to inference.

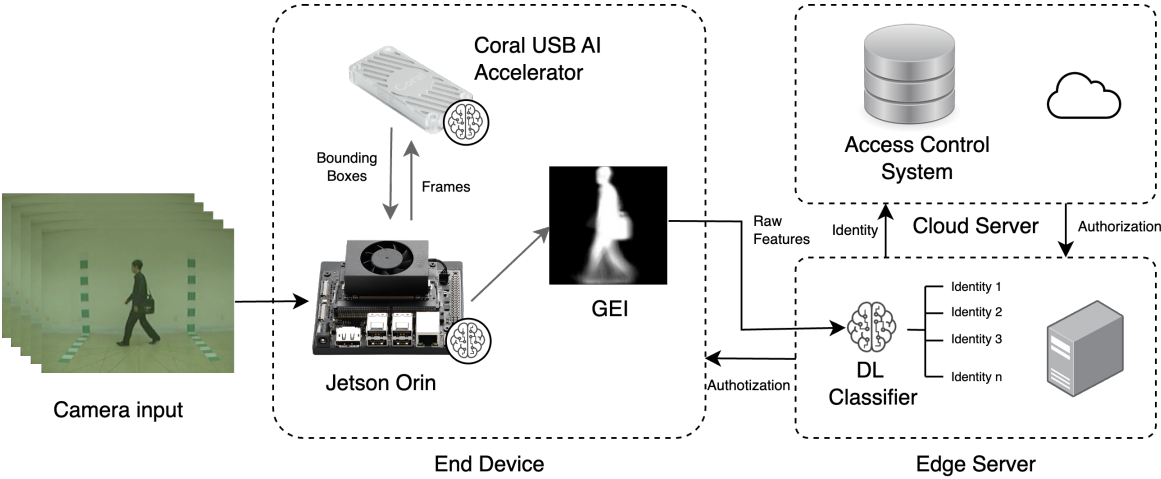


Figure 7. A large-scale scalable framework to support gait recognition computations in a distributed manner. This framework would incorporate multiple nodes, including an Edge Server and End Devices, to handle data acquisition, detection, segmentation, and classification. In scenarios where a node’s capabilities are exceeded, a supporting cloud server would be helpful. By leveraging edge components, the framework could enable more feasible real-time computation.

6.2. End Device-Based

End device-based deployment refers to a system architecture where the entire processing pipeline, from data capture to inference, runs directly on the end device. This approach significantly enhances user privacy and reduces latency, which are crucial factors for real-world applications. Since end devices typically have limited processing power and memory, this type of deployment often relies on model compression techniques to reduce a model’s size and computational requirements. Several

technologies are available to enable this, including hardware peripherals (e.g., microcontrollers and microcomputers powered by CPU, GPU, and TPU chips) and compression tools (such as TensorRT, OpenVINO, and ONNX), discussed further in Section 7.

DL-based systems deployed directly on end devices have found applications in diverse fields, ranging from medical devices that use infrared lighting to locate veins [136] and indoor monitoring for potential airborne disease infection [137], to unmanned surface vehicles for aquatic weed removal [138]. In the realm of CVGR systems, a significant contribution was made by Tiñini et al. [128], who proposed one of the earliest CVGR systems to be fully embedded within a Jetson Nano Development Card equipped with an OAK-D camera. Their approach leveraged a pre-trained person detection model, MobileNetv2, deployed on the OAK-D device, with subsequent representation extraction using a U-Net and inference performed on the Jetson Nano with a CNN. Their proposed system achieved an impressive inference time of 35.8 milliseconds per gait representation recognition.

6.3. Edge Server + End Devices

When deploying the entire CVGR pipeline on an end device is impractical due to limited memory and energy resources, partitioning a computer vision pipeline or model is possible. This involves keeping some parts on an edge server and deploying others on end devices. Factors like network latency, memory footprint, and energy consumption can help determine the optimal number of model splits and the execution location of each part. However, the final deployment strategy should ideally be based on experimentation, empirical criteria, and actual performance evaluation. Neurosurgeon, a scheduling strategy presented in [139], offers a compelling proposal to automate the decisions of where to split a model and how to deploy its portions. Another example of this hybrid approach is found in [140], where armed robbery detection is achieved by combining YOLO-based weapon detectors on end devices with a CNN deployed on an edge server. The edge server then sends notifications to homeowners only if anomalies are confirmed by both the end devices and itself, thus reducing the false positive rate.

6.4. Edge Server + Cloud Server

Also known as edge-cloud, this approach aims to balance computing tasks between cloud servers and edge servers when the latter's capabilities are insufficient. This is particularly valuable now that every data instance is crucial and can contribute to the generalisation of DL models. For example, an edge server can collect and process individual samples sent from end devices, with this data being synchronised with a cloud server once the edge server reaches its memory capacity. This hybrid approach can address the limitations of traditional cloud-based approaches for video analysis, which often suffer from prohibitive bandwidth consumption and high response latency. For example, Han et al. [141] proposed ECCVideo, an edge-cloud collaborative video analysis system that focuses on low-latency applications by efficiently utilising resources at both the edge and in the cloud. This approach has proven successful in scenarios where the capabilities of edge servers alone are not sufficient.

7. Model Optimisation for Edge Computing

The core principle of edge computing is to enable end devices to process the information they gather and respond accordingly, with minimal delay. However, some state-of-the-art applications utilise complex and computationally expensive models, such as those based on DL. Therefore, rather than offloading the computation to an external server, alternative approaches such as model optimisation are essential for deploying DL models on edge devices. This section will cover the following optimisation methods: model design, pruning, quantisation, and knowledge distillation. For a more in-depth exploration of inference acceleration methods for edge computing, see [142] and [143].

7.1. Model Design or Selection

In edge computing, researchers often prioritise designing DNN models with fewer parameters. The goal is to minimise memory usage and execution delays while maintaining high accuracy. This is particularly important when designing DNN models for devices with limited computational capacity, such as edge devices [142]. Iandola et al. [144] proposed SqueezeNet to offer advantages like reduced communication across devices during training, less bandwidth consumption during model export, and the feasibility of deployment on embedded hardware, all while preserving the accuracy of AlexNet-level networks. The authors compressed SqueezeNet to less than 0.5 Megabytes (MB), resulting in a network with 50 times fewer parameters than AlexNet. Similarly, Fang et al. [145] proposed Tinier-YOLO to minimise model size and improve both detection accuracy and real-time performance. Tinier-YOLO resulted in a model size of 8.9 MB, four times smaller than Tiny-YOLOv3. It achieved a mean Average Precision of 65.7% and reached 25 FPS real-time performance on a Jetson TX1.

7.2. Pruning

The pruning technique involves setting redundant or less important neuron parameters to zero without significantly impacting the accuracy of the results [146]. This strategy is widely used in DL model optimisation to reduce memory size and computational complexity. Recently, Yu et al. [147] proposed EasiEdge, a global pruning method to compress and accelerate DNNs for efficient edge computing. By adopting an Alternating Direction Method of Multipliers (ADMM) to decouple the pruning problem into a performance-improving problem and a global pruning sub-problem, their method achieved remarkable results. When pruning 80% of filters in VGG-16, the accuracy dropped by only 0.22%, and the GPU latency on a Jetson TX2 decreased to 0.19ms. Furthermore, Woo et al. [148] presented “zero-keep filter pruning” for creating energy-efficient DNNs. This strategy replaces small values with zero and prunes the filter with the least number of zeros, thus maximising the number of zero elements in filters. Increasing the number of zero elements is expected to reduce the significant power/energy consumption associated with multiplication calculations through the use of a zero-skip multiplier.

7.3. Quantisation

Fundamentally, quantisation is the process of approximating continuous data, whether a single value or a series, using a set of integers. In the context of edge computing, parameter quantisation converts existing floating-point DNN parameters to lower-bit values. This eliminates expensive floating-point operations, thereby reducing the model’s computational weight and cost [142]. In this context, Zebin et al. [149] presented in 2019 a CNN model for classifying five activities of daily living. Their model used raw accelerometer and gyroscope data from a wearable sensor as input. Through model optimisation, they quantised weights per channel and activations per layer to 8-bit accuracy after training. This resulted in a CNN architecture with classification accuracy within 2% of floating-point networks. Notably, weight quantisation was responsible for nearly all of the size and runtime reductions in the improved model. Later, in 2021, Wardana et al. [132] used four distinct post-quantisation approaches provided by the TensorFlow Lite framework to develop an efficient model for edge devices: dynamic range quantisation, float16 quantisation, integer quantisation with float, and complete integer quantisation. Among these, full-integer quantisation yielded the lowest execution time, with latencies of 2.19 seconds and 4.73 seconds for Raspberry Pi 4B and Raspberry Pi 3B+, respectively.

7.4. Knowledge Distillation

Knowledge distillation is a technique where a smaller, simpler model (the “student”) is trained to mimic the behaviour of a larger, more complex model (the “teacher”) [150]. The objective is for the student model to approximate the function learned by the teacher model, enabling powerful models to be deployed on resource-constrained devices where running the full-sized teacher model would

be impractical. Moreover, smaller models are faster and require less memory, making them suitable for real-time applications. Common distillation strategies include Logits Distillation (the student learns to mimic the output probabilities of the teacher), Feature Distillation (the student learns to mimic intermediate layer activations of the teacher), and Relationship Distillation (the student learns to mimic the relationships between different neurons or layers inside the teacher). Recently, Li et al. [4] proposed applying knowledge distillation to CVGR systems, introducing Multi-teacher Joint Knowledge Distillation (MJKD) to address challenges in cross-view gait recognition. Specifically, MJKD uses multiple teacher models to train a lightweight student model, improving its ability to extract gait features and recognise individuals. Experiments on the CASIA-B dataset [23] demonstrated MJKD's effectiveness, achieving 98.24% accuracy while reducing model complexity and computational cost.

8. Gait Recognition on the Edge

Gait recognition systems based on edge computing are a relatively new field, with few academic works addressing this combination (as listed in Table 6). In contrast, multiple companies now offer gait recognition using a cloud-server architecture commercially [151,152]. In the realm of CVGR systems, Tinini et al. [128] proposed a pioneering advance combining edge computing and model-free gait recognition, laying the groundwork for developing more robust and scalable real-time CVGR systems. Nevertheless, to consider a real-life application of this technology and achieve beneficial applications for society, much work remains to be done regarding the correct classification rate (CCR), different viewing angles, and robustness against intra-class and environment variations.

In 2022, Ruiz-Barroso et al. [153] conducted a comprehensive evaluation of hardware and software optimisations for deploying gait recognition models on embedded systems. They analysed the energy consumption and performance of three distinct models on two popular embedded platforms: the NVIDIA Jetson Nano 4 GB and the NVIDIA Jetson AGX Xavier. The study found that targeted hardware and software optimisations (specifically, pruning and quantisation techniques) significantly enhanced execution time, achieving a 4.2x speedup compared to the baseline models. Moreover, these optimisations resulted in a substantial improvement in energy consumption, with the deployed models consuming up to 5.4 times less energy compared to the baseline deployment.

Recently, in 2023, Ma and Liu [33] introduced a new method for gait recognition using millimeter-wave radar. Specifically, they proposed a spatial-temporal network that processes 4D radar point cloud videos to capture the motion dynamics of walking people. The network employs PointNet to extract spatial features from each frame, and then models the spatial-temporal information using a Transformer layer. Experimental results demonstrate that the proposed network outperforms similar methods based on 4D radar reconstruction. Importantly, its ability to function under low-resource setups opens the door for real-time surveillance systems using 4D gait representations.

Deploying a CVGR system in real-world edge environments presents several significant challenges, which can be broadly categorised into three areas: variations in gait, environmental challenges, and security challenges. In the following sections, we will describe each of these challenge groups and discuss potential solutions proposed in the literature. Our goal is to simplify the decision-making process for those seeking to bring this biometric technology into real-world applications.

Table 6. Summary of recent gait recognition systems for person identification approached with edge computing.

Year & Reference	Sensor	Gait Representation/Processing	End device	Gait Feature Classification Model	Results
2024 [13]	Accelerometer/Gyro	2D FFT	B. Akida/Arduino	CNN	0.97 accuracy
2023 [33]	FMCW-MIMO Radar	4D Point cloud videos	Laptop	Transformer	0.91 accuracy
2022 [153]	RGB Camera	GEI	Jetson Nano/AGX	2D/3D CNN	↓ 4.2 secs per inference
2022 [154]	Inertial sensor	Gait Feature Map	Cellphone	CNN+LSTM	0.97 accuracy
2021 [128]	OAK-D Camera	MGEI & PCA	Jetson Nano	MobileNetv2, CNN	0.96 accuracy
2021 [11]	Floor sensors	Spatiotemporal signals	Raspberry Pi	CNN	0.91 f-score

8.1. Variations in Gait

Changes in clothing, footwear, carrying objects, walking speed and emotional state can alter a person's gait pattern, making consistent recognition challenging. For example, a heavy coat or backpack can obscure the body's silhouette and impact limb movement [1]. To address these challenges, researchers are exploring various strategies, such as:

- Developing robust feature extraction techniques and improving feature representations, such as utilising partitioned gait energy images [85] or adopting DL architectures [98], to reduce sensitivity to such variations.
- Employing view transformation models to normalise gait patterns across different perspectives [97].
- Creating synthetic gait data with controlled covariate gait variations to enhance the training of gait recognition models [105,155].
- Implementing federated learning [156] - a distributed approach to data collection, model training and weight sharing - for multi-view gait recognition.

Similarly, variations in viewpoint pose a common challenge for CVGR systems deployed on edge devices. Since these devices can be installed at different heights or with varying camera angles, even minor changes in setup can lead to perspective distortions that hinder consistent comparison of gait features. Notable works addressing this issue include studies on gait recognition under view-invariant scenarios [110], and the use of GANs for covariate-aware gait synthesis [155]. Such techniques can be applied on cloud premises to distribute later better-trained models on the edge.

8.2. Environmental Challenges

Environmental variations between laboratory and real-world settings can encompass a wide range of challenges that significantly impact the performance of CVGR systems. For instance, basic changes, such as walking surfaces (e.g., uneven terrain, stairs), can make it difficult to extract reliable gait features. Furthermore, different lighting and shadows are common issues for CVGR systems under real-world conditions. In such cases, applying a Gamma Filter, Contrast Stretching, or other image processing techniques can enhance the input image. Another frequent problem is the presence of occlusions (including people walking behind objects). To develop robust CVGR systems, gait recognition under occlusion scenarios has been developed for several years. For example, researchers in [157] recently proposed the combination of a VGG-16 model to detect occlusions and an LSTM model to reconstruct occluded frames in the sequence. In 2023, Xu et al. [6] introduced a technique for aligning silhouettes, which involves flexibly estimating the spatial extent of elements causing occlusion.

Moreover, real-world deployments often involve cameras with limited resolution, frame rate, or placement, leading to noisy and incomplete gait data. Addressing these real-world complexities requires developing robust algorithms and acquiring diverse training data that reflects such variability, as proposed in [105,155,158]. These advancements aim to improve the adaptability and robustness of gait recognition systems in real-world setups and under varying conditions.

8.3. Security Challenges

As emerging biometric technologies continue to gain widespread adoption, so too do the methods employed to breach their security and gain access to sensitive user information for various purposes. Automated biometrics, often utilised in authentication and surveillance applications, are not immune to these challenges. Consequently, with the rise of various methods to compromise these systems, a more profound security analysis is imperative for the secure deployment of gait recognition systems at the edge. The following subsections explore three of the most frequent approaches to breaking CVGR systems and their potential countermeasures.

8.3.1. Person Physical Impersonation

From the early 2010s, despite initial promising results, silhouette-based gait recognition systems raised concerns within the scientific community regarding their effectiveness as a biometric method. This prompted several investigations to test their vulnerabilities by having subjects mimic the walk style or clothing of others. For example, in 2012, Hadid et al. [159] published the first analysis on the spoofing susceptibility of such systems. They focused on experimenting with clothing and shape impersonation for two gait recognition systems, UOULU and USOU. These systems were originally designed to extract and process 2D and 3D visual data of walkers from a fixed viewpoint in a tunnel. The researchers then had 22 subjects walk through the same tunnel with similar clothes and body shapes, subsequently calculating the similarity between the recognised gaits. They concluded that while silhouette-based recognition systems were indeed susceptible to spoofing, they were not as easily compromised as face or fingerprint recognition systems.

In 2015, Hadid et al. [160] proposed countermeasures to avoid physical spoofing. Their main countermeasure divides each silhouette within a frame sequence into multiple horizontal portions and computes Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) for each portion. These portions are then combined into a single histogram vector, which is compared to determine if the portions belong to the same person.

8.3.2. Spoofing via Synthetic Data Generation

Recent advancements in DL-based synthetic data generation techniques, particularly autoencoders and GANs, have enabled the automated creation of fake human data such as face images, voice audio, and motion videos. In this context, Jia et al. [161] explored a spoofing method to compromise gait recognition systems. They proposed a GAN-based approach to render target videos on a given scene image from walking sequences. The focus was on generating realistic synthetic videos with good visual effects while also transmitting sufficient information to deceive gait recognition systems. Their experiments with GaitSet [93] and a CNN-Gait-based approach validated the method's effectiveness. Similarly, Hirose et al. [162] proposed a method to transform gait genuine clones (GGCs) from actual walking people into gait silhouette clones (GSCs) using a multiple autoencoder approach. To address potential security concerns, they also presented a supervised learning method to discriminate between genuine and synthetically generated gait silhouettes.

8.3.3. Silhouette Poisoning for Misclassification

Similar to other DL-based biometric systems, gait recognition is also vulnerable to adversarial attacks. These attacks involve adding noise, in the form of patches or pixels, to a gait video, causing the deep learning models to misclassify the sample. For example, in a gait recognition-based entrance security system compromised by such noise, the model might mistakenly recognise an unknown person as someone authorised to enter. To address this, Maqsood et al. [163] recently presented an approach to locate strategic sections in Gait Energy Images for adding imperceptible noise using Grey Wolf Optimisation [164]. Their findings revealed that high-frequency textures are the most effective locations for such noise.

To enhance privacy in gait recognition, researchers in [165] proposed a privacy-preserving approach to anonymise human gait in videos by deforming silhouettes. They demonstrated that publicly available videos of people walking could be used to identify them without their knowledge and proposed deforming extracted gait silhouettes and replacing the original visual information in the videos with the modified ones. This resulted in a significant reduction in recognition accuracy from 100% to 1.57%, highlighting the approach's potential to enhance privacy for individuals, albeit at the cost of model accuracy.

9. Discussion and Future Work

9.1. Trends in Computer Vision-Based Gait Recognition Systems

Advances in CVGR systems are showing promising results, increasing the trustworthiness of this biometric technology. Moreover, the widespread use of closed-circuit television (CCTV) and low-cost cameras presents significant potential for integrating CVGR systems to enhance surveillance in public and private settings. However, the pursuit of higher precision has led to increasingly complex and resource-intensive models, often requiring deployment in high-performance computing environments. This limits the technology's accessibility in scenarios with unstable internet connections or limited server capacity. Edge computing architectures offer a promising solution by distributing inference capabilities, thus addressing this limitations. As discussed in Section 7, several studies have demonstrated the effectiveness of edge computing architectures, highlighting opportunities to apply these to CVGR systems.

In terms of models, distance-based gait sequence classifiers remain the most straightforward and resource-efficient approach. However, machine learning becomes crucial when generalisation is critical for a CVGR project. Figures 1a and 1b show a trend towards employing DL models in CVGR systems, with CNNs, RNN, LSTM, GAN, CapsNets, and Transformers being the most frequent. Despite remarkable advancements in DL-based CVGR systems, their real-world application in uncontrolled environments remains challenging, although recent works and datasets, such as [14] and [16], demonstrate promising results. For example, recent results obtained by DL models with challenging gait sequence variations, such as GPGait [7] and SkeletonGait [43], highlight opportunities for model-based approaches. On top of that, novel gait recognition technologies such as depth sensing, stereo vision, 3d laser scene reconstruction, pose estimation, optical flow, and graph-based models hold promising research threads for future investigations.

Newer technologies have yet to overcome challenges such as variations in illumination. Changes in lighting due to time of day, weather, and shadows can distort silhouette- or appearance-based features, affecting gait extraction. Another challenge is occlusions, where objects, other people, or the environment obscure body parts, leading to incomplete gait data. Viewpoint variation is also critical. Unlike controlled indoor setups with consistent camera angles, outdoor environments often involve dynamic viewpoints, making it difficult to match gait features across different perspectives when implementing CVGR systems [7].

9.2. Edge Computing for CVGR Systems

Given the computational complexity of gait data acquisition, representation, dimensionality reduction, and recognition, deploying a DL-based workflow on a single device can be demanding. However, a distributed edge computing approach offers a solution by enabling task distribution across multiple devices. As detailed in Section 6, several architectures for distributing inference are currently used in academia and industry. These architectures are applicable to CVGR systems and can be further researched for optimisation when implementing CVGR systems coupled with new sensing modalities (3D laser, stereo vision) or feature representations (skeleton maps, point clouds, graph-based models), which still require high-capacity servers. Nevertheless, the data- and resource-intensive nature of deep learning necessitates larger datasets and more complex pipelines, making it challenging to embed these pipelines in resource-constrained environments.

This article revealed several promising technologies for developing CVGR systems suitable for edge devices. For example, 2D imaging currently offers advantages over 3D scene reconstruction technologies (depth sensing, stereo vision, 3D lasers). Most model-free gait feature representations discussed in Section 3.2 are advantageous for deployment in edge computing architectures, while the growing precision of model-based approaches holds promising opportunities for future investigations. Furthermore, the nature of graph-based models shows promise for embedding in low-resource devices; however, they currently require further optimisation methods (detailed in Section 7). Finally, dimen-

sionality reduction techniques remain relevant due to the high dimensionality of gait data, especially when using multi-modal and multi-camera approaches for gait recognition.

Moving on, the increasing availability of powerful Edge AI development kits, such as the NVIDIA Jetson family (including the Jetson Nano, Jetson Xavier NX, and Jetson AGX Orin), Google Coral Dev Board, Intel Edge AI Box, Raspberry Pi, Qualcomm Snapdragon AI Development Kit, Luxonis OAK-D, Hailo-8 AI processor, Xilinx Kria K26 SOM, Edge Impulse Kits, and Khadas VIM, is opening opportunities to embed more complex pipelines at the edge. These kits enable parallel computing through dedicated chips (GPUs, VPUs, TPUs) or model compression and optimisation techniques (e.g., NVIDIA TensorRT, Intel OpenVINO, TensorFlow Lite). Table 6 showcases examples of CVGR systems designed for resource-constrained devices, highlighting a growing research trend in combining CVGR and edge computing. Also, Table 5 shows multiple successful cases where computer vision systems were combined with edge computing.

In summary, future CVGR systems should explore modular frameworks that leverage edge computing. Distributing gait recognition tasks across multiple edge devices can significantly improve inference speed while maintaining accuracy and enhancing system robustness by capturing gait data from diverse viewpoints. This approach also lays the foundation for improved multi-modal or multi-view frameworks, where different end devices capture data from the same subject for subsequent recognition, offering immense potential for advancing both gait recognition and computer vision.

10. Conclusions

Gait recognition is achieving increasingly compelling results every year thanks to advances in deep learning architectures. However, most recent approaches still rely on sending video of gait sequences to a central server for processing, which necessitates an internet connection and can be costly. In contrast, edge computing proposes collecting and processing data closer to the source on local devices, offering faster response times and enabling real-time monitoring. This trend, coupled with the availability of powerful mini-computers, makes it possible to run advanced computer vision-based gait recognition (CVGR) systems directly on edge devices where the gait data is captured.

This paper reviews different edge computing architectures for computer vision-based gait recognition, discusses their design, deployment, and explores the challenges and future possibilities of integrating these technologies. The paper emphasises methods suitable for low-resource devices and identifies areas where further research is needed for optimisation. Importantly, the review highlights a significant research gap: the limited exploration of combining CVGR and edge computing. This integration could substantially improve gait recognition implementations, making them more robust and suitable for real-world applications, including real-time surveillance and smart authentication systems.

Funding: This research received no external funding.

Acknowledgments: We appreciate the support of Christian Conchari and Guillermo Sahonero from Universidad Católica Boliviana “San Pablo” for being involved in the project and helping us define the foundation of the research gap.

Conflicts of Interest: The author declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CCR	Correct Classification Rate
CCTV	Closed-Circuit Television
CNN	Convolutional Neural Network
CVGR	Computer Vision-based Gait Recognition
DL	Deep Learning
DR	Dimensionality Reduction
GAN	Generative Adversarial Networks

GPU	Graphics Processing Unit
LDA	Linear Discriminant Analysis
LiDAR	Light Detection and Ranging
PCA	Principal Component Analysis
ML	Machine Learning
SGF	Stereo Gait Feature
TPU	Tensor Processing Unit
VPU	Virtual Processing Unit

References

1. Singh, J.P.; Jain, S.; Arora, S.; Singh, U.P. Vision-based gait recognition: A survey. *IEEE Access* **2018**, *6*, 70497–70527.
2. Marsico, M.D.; Mecca, A. A Survey on Gait Recognition via Wearable Sensors. *ACM Comput. Surv.* **2019**, *52*, 1–39.
3. Wan, C.; Wang, L.; Phoha, V.V. A Survey on Gait Recognition. *ACM Comput. Surv.* **2018**, *51*, 1–35.
4. Li, R.; Yun, L.; Zhang, M.; Yang, Y.; Cheng, F. Cross-View Gait Recognition Method Based on Multi-Teacher Joint Knowledge Distillation. *Sensors* **2023**, *23*, 9289.
5. Zheng, J.; Liu, X.; Liu, W.; He, L.; Yan, C.; Mei, T. Gait recognition in the wild with dense 3d representations and a benchmark. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20228–20237.
6. Xu, C.; Tsuji, S.; Makihara, Y.; Li, X.; Yagi, Y. Occluded Gait Recognition via Silhouette Registration Guided by Automated Occlusion Degree Estimation. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 3199–3209.
7. Fu, Y.; Meng, S.; Hou, S.; Hu, X.; Huang, Y. Gpgait: Generalized pose-based gait recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19595–19604.
8. Shen, C.; Yu, S.; Wang, J.; Huang, G.Q.; Wang, L. A comprehensive survey on deep gait recognition: algorithms, datasets and challenges. *arXiv preprint arXiv:2206.13732* **2022**.
9. Güner Şahan, P.; Şahin, S.; Kaya Gülağız, F. A survey of appearance-based approaches for human gait recognition: techniques, challenges, and future directions. *The Journal of Supercomputing* **2024**, pp. 1–38.
10. Ahmed, D.M.; Mahmood, B.S. Survey: gait recognition based on image energy. *2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC)*. IEEE, 2023, pp. 51–56.
11. Alharthi, A.S.; Casson, A.J.; Ozanyan, K.B. Spatiotemporal Analysis by Deep Learning of Gait Signatures From Floor Sensors. *IEEE Sens. J.* **2021**, *21*, 16904–16914.
12. Li, J.; Wang, Z.; Zhao, Z.; Jin, Y.; Yin, J.; Huang, S.L.; Wang, J. TriboGait: A deep learning enabled triboelectric gait sensor system for human activity recognition and individual identification. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 643–648.
13. Venkatachalam, S.; Nair, H.; Vellaisamy, P.; Zhou, Y.; Youssfi, Z.; Shen, J.P. Realtime Person Identification via Gait Analysis. *arXiv preprint arXiv:2404.15312* **2024**.
14. Song, C.; Huang, Y.; Wang, W.; Wang, L. CASIA-E: a large comprehensive dataset for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2022**, *45*, 2801–2815.
15. Mu, Z.; Castro, F.M.; Marin-Jimenez, M.J.; Guil, N.; Li, Y.R.; Yu, S. Resgait: The real-scene gait dataset. *2021 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 2021, pp. 1–8.
16. Zhu, Z.; Guo, X.; Yang, T.; Huang, J.; Deng, J.; Huang, G.; Du, D.; Lu, J.; Zhou, J. Gait recognition in the wild: A benchmark. *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 14789–14799.
17. Dou, H.; Zhang, W.; Zhang, P.; Zhao, Y.; Li, S.; Qin, Z.; Wu, F.; Dong, L.; Li, X. Versatilegait: a large-scale synthetic gait dataset with fine-grained attributes and complicated scenarios. *arXiv preprint arXiv:2101.01394* **2021**.
18. An, W.; Yu, S.; Makihara, Y.; Wu, X.; Xu, C.; Yu, Y.; Liao, R.; Yagi, Y. Performance evaluation of model-based gait on multi-view very large population database with pose sequences. *IEEE transactions on biometrics, behavior, and identity science* **2020**, *2*, 421–430.

19. Iwashita, Y.; Ogawara, K.; Kurazume, R. Identification of people walking along curved trajectories. *Pattern Recognition Letters* **2014**, *48*, 60–69.
20. Samangoeei, S.; Bustard, J.; Nixon, M.; Carter, J. On acquisition and analysis of a dataset comprising of gait, ear and semantic data. *Multibiometrics for Human Identification* **2011**, pp. 277–301.
21. Sivapalan, S.; Chen, D.; Denman, S.; Sridharan, S.; Fookes, C. Gait energy volumes and frontal gait recognition using depth images. 2011 International Joint Conference on Biometrics (IJCB). IEEE, 2011, pp. 1–6.
22. Tan, D.; Huang, K.; Yu, S.; Tan, T. Efficient night gait recognition based on template matching. 18th international conference on pattern recognition (ICPR'06). IEEE, 2006, Vol. 3, pp. 1000–1003.
23. Yu, S.; Tan, D.; Tan, T. A Framework for Evaluating the Effect of View Angle, Clothing and Carrying Condition on Gait Recognition. 18th International Conference on Pattern Recognition (ICPR'06), 2006, Vol. 4, pp. 441–444.
24. Wang, L.; Tan, T.; Ning, H.; Hu, W. Silhouette analysis-based gait recognition for human identification. *IEEE transactions on pattern analysis and machine intelligence* **2003**, *25*, 1505–1518.
25. Huang, P.S.; Harris, C.J.; Nixon, M.S. Recognising humans by gait via parametric canonical space. *Artificial Intelligence in Engineering* **1999**, *13*, 359–366.
26. Sadeghzadehyazdi, N.; Batabyal, T.; Acton, S.T. Modeling spatiotemporal patterns of gait anomaly with a CNN-LSTM deep neural network. *Expert Syst. Appl.* **2021**, *185*, 115582.
27. Chattopadhyay, P.; Roy, A.; Sural, S.; Mukhopadhyay, J. Pose Depth Volume extraction from RGB-D streams for frontal gait recognition. *J. Vis. Commun. Image Represent.* **2014**, *25*, 53–63.
28. Bari, A.S.M.H.; Gavrilova, M.L. Artificial Neural Network Based Gait Recognition Using Kinect Sensor. *IEEE Access* **2019**, *7*, 162708–162722.
29. Albert, J.A.; Owolabi, V.; Gebel, A.; Brahms, C.M.; Granacher, U.; Arnrich, B. Evaluation of the Pose Tracking Performance of the Azure Kinect and Kinect v2 for Gait Analysis in Comparison with a Gold Standard: A Pilot Study. *Sensors* **2020**, *20*.
30. Wang, J.; She, M.; Nahavandi, S.; Kouzani, A. A Review of Vision-Based Gait Recognition Methods for Human Identification. 2010 International Conference on Digital Image Computing: Techniques and Applications, 2010, pp. 320–327.
31. Liao, R.; Yu, S.; An, W.; Huang, Y. A model-based gait recognition method with body pose and human prior knowledge. *Pattern Recognit.* **2020**, *98*.
32. Shen, C.; Fan, C.; Wu, W.; Wang, R.; Huang, G.Q.; Yu, S. Lidargait: Benchmarking 3d gait recognition with point clouds. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1054–1063.
33. Ma, C.; Liu, Z. A Novel Spatial–Temporal Network for Gait Recognition Using Millimeter-Wave Radar Point Cloud Videos. *Electronics* **2023**, *12*, 4785.
34. Li, Y.; Zhang, P.; Zhang, Y.; Miyazaki, K. Gait analysis using stereo camera in daily environment. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019, pp. 1471–1475.
35. Liu, H.; Cao, Y.; Wang, Z. Automatic gait recognition from a distance. 2010 Chinese Control and Decision Conference. IEEE, 2010, pp. 2777–2782.
36. Kok, K.Y.; Rajendran, P. A review on stereo vision algorithm: Challenges and solutions. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)* **2019**, *13*, 112–128.
37. Dai, J.; Li, Q.; Wang, H.; Liu, L. Understanding images of surveillance devices in the wild. *Knowledge-Based Systems* **2024**, *284*, 111226.
38. Stenum, J.; Rossi, C.; Roemmich, R.T. Two-dimensional video-based analysis of human gait using pose estimation. *PLoS computational biology* **2021**, *17*, e1008935.
39. Hii, C.S.T.; Gan, K.B.; Zainal, N.; Mohamed Ibrahim, N.; Azmin, S.; Mat Desa, S.H.; van de Warrenburg, B.; You, H.W. Automated gait analysis based on a marker-free pose estimation model. *Sensors* **2023**, *23*, 6489.
40. Teepe, T.; Khan, A.; Gilg, J.; Herzog, F.; Hörmann, S.; Rigoll, G. Gaitgraph: Graph convolutional network for skeleton-based gait recognition. 2021 IEEE international conference on image processing (ICIP). IEEE, 2021, pp. 2314–2318.

41. Teepe, T.; Gilg, J.; Herzog, F.; Hörmann, S.; Rigoll, G. Towards a deeper understanding of skeleton-based gait recognition. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1569–1577.
42. Konz, L.; Hill, A.; Banaei-Kashani, F. ST-DeepGait: A spatiotemporal deep learning model for human gait recognition. *Sensors* **2022**, *22*, 8075.
43. Fan, C.; Ma, J.; Jin, D.; Shen, C.; Yu, S. SkeletonGait: Gait Recognition Using Skeleton Maps. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, Vol. 38, pp. 1662–1669.
44. Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. Multi-view large population gait dataset and its performance evaluation for cross-view gait recognition. *IPSN transactions on Computer Vision and Applications* **2018**, *10*, 1–14.
45. Bobick, A.F.; Davis, J.W. The recognition of human movement using temporal templates. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 257–267.
46. Lam, T.H.W.; Lee, R.S.T. A New Representation for Human Gait Recognition: Motion Silhouettes Image (MSI). *Advances in Biometrics*. Springer Berlin Heidelberg, 2005, pp. 612–618.
47. Han, J.; Bhanu, B. Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 316–322.
48. Liu, J.; Zheng, N. Gait history image: A novel temporal template for gait recognition. 2007 IEEE international conference on multimedia and expo. IEEE Computer Society, 2007, pp. 663–666.
49. Zebhi, S.; AlModarresi, S.M.T.; Abootalebi, V. Human activity recognition using pre-trained network with informative templates. *International Journal of Machine Learning and Cybernetics* **2021**, *12*, 3449–3461.
50. Wang, M.; Guo, X.; Lin, B.; Yang, T.; Zhu, Z.; Li, L.; Zhang, S.; Yu, X. DyGait: Exploiting dynamic representations for high-performance gait recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13424–13433.
51. Bashir, K.; Xiang, T.; Gong, S. Gait recognition using Gait Entropy Image. 3rd International Conference on Imaging for Crime Detection and Prevention (ICDP 2009), 2009, pp. 1–6.
52. Lam, T.H.; Cheung, K.H.; Liu, J.N. Gait flow image: A silhouette-based gait representation for human identification. *Pattern recognition* **2011**, *44*, 973–987.
53. Ye, H.; Sun, T.; Xu, K. Gait Recognition Based on Gait Optical Flow Network with Inherent Feature Pyramid. *Applied Sciences* **2023**, *13*, 10975.
54. Chen, C.; Liang, J.; Zhao, H.; Hu, H.; Tian, J. Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recognit. Lett.* **2009**, *30*, 977–984.
55. Wang, C.; Zhang, J.; Pu, J.; Yuan, X.; Wang, L. Chrono-gait image: A novel temporal template for gait recognition. *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision*, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part I 11. Springer, 2010, pp. 257–270.
56. Zhang, E.; Zhao, Y.; Xiong, W. Active energy image plus 2DLPP for gait recognition. *Signal Processing* **2010**, *90*, 2295–2302.
57. He, Y.; Zhang, J.; Shan, H.; Wang, L. Multi-Task GANs for view-specific feature learning in gait recognition. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 102–113.
58. Wang, X.; Yan, W.Q. Human Gait Recognition Based on Frame-by-Frame Gait Energy Images and Convolutional Long Short-Term Memory. *Int. J. Neural Syst.* **2020**, *30*, 1950027.
59. Wang, K.; Liu, L.; Lee, Y.; Ding, X.; Lin, J. Nonstandard Periodic Gait Energy Image for Gait Recognition and Data Augmentation. *Pattern Recognition and Computer Vision*. Springer International Publishing, 2019, pp. 197–208.
60. Bharadwaj, S.V.; Chanda, K. Person Re-Identification by Analyzing Dynamic Variations in Gait Sequences. *Evolving Technologies for Computing, Communication and Smart World*. Springer Singapore, 2021, pp. 399–410.
61. Luo, J.; Zhang, J.; Zi, C.; Niu, Y.; Tian, H.; Xiu, C. Gait recognition using GEI and AFDEI. *International Journal of Optics* **2015**, *2015*, 763908.
62. Thomas, K.; Pushpalatha, K. A comparative study of the performance of gait recognition using gait energy image and shannon's entropy image with CNN. *Data Science and Security: Proceedings of IDSCS 2021*. Springer, 2021, pp. 191–202.

63. Filipi Gonçalves dos Santos, C.; Oliveira, D.d.S.; A. Passos, L.; Gonçalves Pires, R.; Felipe Silva Santos, D.; Pascotti Valem, L.; P. Moreira, T.; Cleison S. Santana, M.; Roder, M.; Paulo Papa, J.; Colombo, D. Gait Recognition Based on Deep Learning: A Survey. *ACM Comput. Surv.* **2022**, *55*, 1–34.
64. Wattanapanich, C.; Wei, H.; Petchkit, W. Investigation of robust gait recognition for different appearances and camera view angles. *Int. J. Elect. Computer Syst. Eng.* **2021**, *11*, 3977–3987.
65. Khalifa, O.O.; Jawed, B.; Bhuiyn, S.S.N. Principal component analysis for human gait recognition system. *Bulletin of Electrical Engineering and Informatics* **2019**, *8*, 569–576.
66. Hasan, B.M.S.; Abdulazeez, A.M. A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *jscdm* **2021**, *2*, 20–30.
67. Gupta, S.K.; Sultaniya, G.M.; Chattopadhyay, P. An Efficient Descriptor for Gait Recognition Using Spatio-Temporal Cues. *Emerging Technology in Modelling and Graphics*. Springer Singapore, 2020, pp. 85–97.
68. Guo, H.; Li, B.; Zhang, Y.; Zhang, Y.; Li, W.; Qiao, F.; Rong, X.; Zhou, S. Gait Recognition Based on the Feature Extraction of Gabor Filter and Linear Discriminant Analysis and Improved Local Coupled Extreme Learning Machine. *Math. Probl. Eng.* **2020**, 2020.
69. Wang, H.; Fan, Y.; Fang, B.; Dai, S. Generalized linear discriminant analysis based on euclidean norm for gait recognition. *International Journal of Machine Learning and Cybernetics* **2018**, *9*, 569–576.
70. Li, W.; Kuo, C.C.J.; Peng, J. Gait recognition via GEI subspace projections and collaborative representation classification. *Neurocomputing* **2018**, *275*, 1932–1945.
71. Yu, S.; Tan, D.; Tan, T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. 18th international conference on pattern recognition (ICPR'06). IEEE, 2006, Vol. 4, pp. 441–444.
72. Honggui, L.; Xingguo, L. Gait analysis using LLE. Proceedings 7th International Conference on Signal Processing, 2004. Proceedings. ICSP'04. 2004. IEEE, 2004, Vol. 2, pp. 1423–1426.
73. Pataky, T.C.; Mu, T.; Bosch, K.; Rosenbaum, D.; Goulermas, J.Y. Gait recognition: highly unique dynamic plantar pressure patterns among 104 individuals. *Journal of The Royal Society Interface* **2012**, *9*, 790–800.
74. Che, L.; Kong, Y. Gait recognition based on DWT and t-SNE. In Proceedings of the Third International Conference on Cyberspace Technology (CCT 2015). IET, 2015, pp. 1423–1426.
75. Huang, X.; Zhu, D.; Wang, H.; Wang, X.; Yang, B.; He, B.; Liu, W.; Feng, B. Context-sensitive temporal feature learning for gait recognition. Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 12909–12918.
76. Pan, Z.; Rust, A.G.; Bolouri, H. Image redundancy reduction for neural network classification using discrete cosine transforms. Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium. ieeexplore.ieee.org, 2000, Vol. 3, pp. 149–154 vol.3.
77. Fan, Z.; Jiang, J.; Weng, S.; He, Z.; Liu, Z. Human gait recognition based on Discrete Cosine Transform and Linear Discriminant Analysis. 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016, pp. 1–6. doi:10.1109/ICSPCC.2016.7753698.
78. Chhatrala, R.; Jadhav, D.V. Multilinear Laplacian discriminant analysis for gait recognition. *IET Comput. Vision* **2017**, *11*, 153–160.
79. Shivani.; Singh, N. Gait Recognition Using DWT and DCT Techniques. Proceedings of International Conference on Communication and Artificial Intelligence. Springer Singapore, 2021, pp. 473–481.
80. Wen, J.; Wang, X. Gait recognition based on sparse linear subspace. *IET Image Proc.* **2021**, *15*, 2761–2769.
81. Liu, Z.; Sarkar, S. Simplest representation yet for gait recognition: averaged silhouette. Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., 2004, Vol. 4, pp. 211–214 Vol.4.
82. Chi, L.; Dai, C.; Yan, J.; Liu, X. An Optimized Algorithm on Multi-view Transform for Gait Recognition. *Communications and Networking*. Springer International Publishing, 2019, pp. 166–174.
83. Das Choudhury, S.; Tjahjadi, T. Robust view-invariant multiscale gait recognition. *Pattern Recognit.* **2015**, *48*, 798–811.
84. Pratama, F.I.; Budianita, A. Optimization of K-Nn Classification In Human Gait Recognition. 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020, pp. 1–5.
85. Premalatha, G.; Chandramani, V. P. Improved gait recognition through gait energy image partitioning. *Comput. Intell.* **2020**, *36*, 1261–1274.

86. Suthaharan, S. Support Vector Machine. In *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning*; Suthaharan, S., Ed.; Springer US: Boston, MA, 2016; pp. 207–235.
87. Wang, X.; Wang, J.; Yan, K. Gait recognition based on Gabor wavelets and (2D)2PCA. *Multimed. Tools Appl.* **2018**, *77*, 12545–12561.
88. Chen, K.; Wu, S.; Li, Z. Gait Recognition Based on GFHI and Combined Hidden Markov Model. 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI). *ieeexplore.ieee.org*, 2020, pp. 287–292.
89. Lu, H.; Plataniotis, K.N.; Venetsanopoulos, A.N. Gait Recognition Through MPCA Plus LDA. 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference. *ieeexplore.ieee.org*, 2006, pp. 1–6.
90. Isaac, E.R.H.P.; Elias, S.; Rajagopalan, S.; Easwarakumar, K.S. Template-based gait authentication through Bayesian thresholding. *IEEE/CAA Journal of Automatica Sinica* **2019**, *6*, 209–219.
91. Venkat, I.; De Wilde, P. Robust Gait Recognition by Learning and Exploiting Sub-gait Characteristics. *Int. J. Comput. Vis.* **2011**, *91*, 7–23.
92. Yousef, R.N.; Khalil, A.T.; Samra, A.S.; Ata, M.M. Model-based and model-free deep features fusion for high performed human gait recognition. *The Journal of Supercomputing* **2023**, *79*, 12815–12852.
93. Chao, H.; He, Y.; Zhang, J.; Feng, J. GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition. *Proceedings of the AAAI Conference on Artificial Intelligence* **2019**, *33*, 8126–8133.
94. Zhang, S.; Wang, Y.; Li, A. Cross-view gait recognition with deep universal linear embeddings. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 9095–9104.
95. Mittal, V.; Bhushan, B. Accelerated computer vision inference with AI on the edge. 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT); Tomar G.S., Ed. Institute of Electrical and Electronics Engineers Inc., 2020, pp. 55–60.
96. Teuwen, J.; Moriakov, N. Chapter 20 - Convolutional neural networks. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, S.K.; Rueckert, D.; Fichtinger, G., Eds.; Academic Press, 2020; pp. 481–501.
97. Min, P.P.; Sayeed, S.; Ong, T.S. Gait Recognition Using Deep Convolutional Features. 2019 7th International Conference on Information and Communication Technology (ICoICT), 2019, pp. 1–5.
98. Ambika, K.; Radhika, K.R. Speed Invariant Human Gait Authentication Based on CNN. Second International Conference on Image Processing and Capsule Networks. Springer International Publishing, 2022, pp. 801–809.
99. Takemura, N.; Makihara, Y.; Muramatsu, D.; Echigo, T.; Yagi, Y. On Input/Output Architectures for Convolutional Neural Network-Based Cross-View Gait Recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *29*, 2708–2719.
100. Lopez Pinaya, W.H.; Vieira, S.; Garcia-Dias, R.; Mechelli, A. Chapter 11 - Autoencoders. In *Machine Learning*; Mechelli, A.; Vieira, S., Eds.; Academic Press, 2020; pp. 193–208.
101. Yu, S.; Wang, Q.; Shen, L.; Huang, Y. View invariant gait recognition using only one uniform model. 2016 23rd International Conference on Pattern Recognition (ICPR), 2016, pp. 889–894.
102. Babaee, M.; Li, L.; Rigoll, G. Person identification from partial gait cycle using fully convolutional neural networks. *Neurocomputing* **2019**, *338*, 116–125.
103. Guo, H.; Ji, Q. Physics-augmented autoencoder for 3d skeleton-based gait recognition. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 19627–19638.
104. Li, F.; Liang, D.; Lian, J.; Liu, Q.; Zhu, H.; Liu, J. Invka: Gait recognition via invertible koopman autoencoder. *arXiv preprint arXiv:2309.14764* **2023**.
105. Yu, S.; Liao, R.; An, W.; Chen, H.; García, E.B.; Huang, Y.; Poh, N. GaitGANv2: Invariant gait feature extraction using generative adversarial networks. *Pattern Recognit.* **2019**, *87*, 179–189.
106. Zhang, P.; Wu, Q.; Xu, J. VT-GAN: View Transformation GAN for Gait Recognition across Views. 2019 international joint conference on neural networks (IJCNN). Institute of Electrical and Electronics Engineers Inc., 2019, Vol. 2019-July, pp. 1–8.
107. Hinton, G.E.; Krizhevsky, A.; Wang, S.D. Transforming Auto-Encoders. *Artificial Neural Networks and Machine Learning – ICANN 2011*. Springer Berlin Heidelberg, 2011, pp. 44–51.

108. Sepas-Moghaddam, A.; Ghorbani, S.; Troje, N.F.; Etemad, A. Gait recognition using multi-scale partial representation transformation with capsules. 2020 25th international conference on pattern recognition (ICPR). Institute of Electrical and Electronics Engineers Inc., 2020, pp. 8045–8052.
109. Zhao, A.; Li, J.; Ahmed, M. SpiderNet: A spiderweb graph neural network for multi-view gait recognition. *Knowledge-Based Systems* **2020**, 206.
110. Xu, Z.; Lu, W.; Zhang, Q.; Yeung, Y.; Chen, X. Gait recognition based on capsule network. *J. Vis. Commun. Image Represent.* **2019**, 59, 159–167.
111. Wu, Y.; Hou, J.; Su, Y.; Wu, C.; Huang, M.; Zhu, Z. Gait Recognition Based on Feedback Weight Capsule Network. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2020, Vol. 1, pp. 155–160.
112. DiPietro, R.; Hager, G.D. Chapter 21 - Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention*; Zhou, S.K.; Rueckert, D.; Fichtinger, G., Eds.; Academic Press, 2020; pp. 503–519.
113. Santos, C.F.G.d.; Oliveira, D.D.S.; Passos, L.A.; Pires, R.G.; Santos, D.F.S.; Valem, L.P.; Moreira, T.P.; Santana, M.C.S.; Roder, M.; Papa, J.P.; others. Gait recognition based on deep learning: a survey. *arXiv preprint arXiv:2201.03323* **2022**.
114. Björnsson, H.H.; Kaldal, J. Exploration and Evaluation of RNN Models on Low-Resource Embedded Devices for Human Activity Recognition, 2023.
115. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern recognition letters* **2019**, 119, 3–11.
116. Sepas-Moghaddam, A.; Etemad, A. View-Invariant Gait Recognition with Attentive Recurrent Learning of Partial Representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **2021**, 3, 124–137.
117. Auten, A.; Tomei, M.; Kumar, R. Hardware acceleration of graph neural networks. 2020 57th ACM/IEEE Design Automation Conference (DAC). IEEE, 2020, pp. 1–6.
118. Lan, T.; Shi, Z.; Wang, K.; Yin, C. Gait Recognition Algorithm based on Spatial-temporal Graph Neural Network. 2022 International Conference on Big Data, Information and Computer Network (BDICN). IEEE, 2022, pp. 55–58.
119. Ma, K.; Fu, Y.; Zheng, D.; Cao, C.; Hu, X.; Huang, Y. Dynamic aggregated network for gait recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22076–22085.
120. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (tog)* **2019**, 38, 1–12.
121. Li, G.; Guo, L.; Zhang, R.; Qian, J.; Gao, S. Transgait: Multimodal-based gait recognition with set transformer. *Applied Intelligence* **2023**, 53, 1535–1547.
122. Wang, X.; Yan, W.Q. Non-local gait feature extraction and human identification. *Multimed. Tools Appl.* **2021**, 80, 6065–6078.
123. Jia, P.; Zhao, Q.; Li, B.; Zhang, J. CJAM: Convolutional Neural Network Joint Attention Mechanism in Gait Recognition. *IEICE Trans. Inf. Syst.* **2021**, E104.D, 1239–1249.
124. Mogan, J.N.; Lee, C.P.; Lim, K.M.; Muthu, K.S. Gait-ViT: Gait Recognition with Vision Transformer. *Sensors* **2022**, 22.
125. Bilal, M.; Jianbiao, H.; Mushtaq, H.; Asim, M.; Ali, G.; ElAffendi, M. GaitSTAR: Spatial–Temporal Attention-Based Feature-Reweighting Architecture for Human Gait Recognition. *Mathematics* **2024**, 12, 2458.
126. Jung, V.J.; Burrello, A.; Scherer, M.; Conti, F.; Benini, L. Optimizing the Deployment of Tiny Transformers on Low-Power MCUs. *arXiv preprint arXiv:2404.02945* **2024**.
127. Dayal, A.; Paluru, N.; Cenkeramaddi, L.R.; J., S.; Yalavarthy, P.K. Design and Implementation of Deep Learning Based Contactless Authentication System Using Hand Gestures. *Electronics* **2021**, 10, 182.
128. Tiñini Alvarez, I.R.; Sahonero-Alvarez, G.; Menacho, C.; Suarez, J. Exploring Edge Computing for Gait Recognition. 2021 4th International Conference on Bio-Engineering for Smart Technologies (BioSMART), 2021, pp. 01–04.
129. Isik, M.; Oldland, M.; Zhou, L. An energy-efficient reconfigurable autoencoder implementation on fpga. Proceedings of SAI Intelligent Systems Conference. Springer, 2023, pp. 212–222.

130. Chen, Z.; Peng, L.; Hu, A.; Fu, H. Generative adversarial network-based rogue device identification using differential constellation trace figure. *EURASIP Journal on Wireless Communications and Networking* **2021**, 2021, 72.
131. Costa, M.; Costa, D.; Gomes, T.; Pinto, S. Shifting capsule networks from the cloud to the deep edge. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2022**, 13, 1–25.
132. Wardana, I.N.K.; Gardner, J.W.; Fahmy, S.A. Optimising Deep Learning at the Edge for Accurate Hourly Air Quality Prediction. *Sensors* **2021**, 21.
133. Jeziorek, K.; Wzorek, P.; Blachut, K.; Pinna, A.; Kryjak, T. Embedded Graph Convolutional Networks for Real-Time Event Data Processing on SoC FPGAs. *arXiv preprint arXiv:2406.07318* **2024**.
134. Wang, Y.; Yang, C.; Lan, S.; Zhu, L.; Zhang, Y. End-edge-cloud collaborative computing for deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials* **2024**.
135. Galanopoulos, A.; Ayala-Romero, J.A.; Leith, D.J.; Iosifidis, G. AutoML for Video Analytics with Edge Computing. *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021, pp. 1–10. doi:10.1109/INFOCOM42981.2021.9488704.
136. Salcedo, E.; Peñaloza, P. Edge AI-Based Vein Detector for Efficient Venipuncture in the Antecubital Fossa. *Mexican International Conference on Artificial Intelligence*. Springer, 2023, pp. 297–314.
137. Rojas, W.; Salcedo, E.; Sahonero, G. ADRAS: airborne disease risk assessment system for closed environments. *Annual International Conference on Information Management and Big Data*. Springer, 2022, pp. 96–112.
138. Salcedo, E.; Uchani, Y.; Mamani, M.; Fernandez, M. Towards Continuous Floating Invasive Plant Removal Using Unmanned Surface Vehicles and Computer Vision. *IEEE Access* **2024**.
139. Kang, Y.; Hauswald, J.; Gao, C.; Rovinski, A.; Mudge, T.; Mars, J.; Tang, L. Neurosurgeon: Collaborative Intelligence Between the Cloud and Mobile Edge. *SIGARCH Comput. Archit. News* **2017**, 45, 615–629. doi:10.1145/3093337.3037698.
140. Fernandez-Testa, S.; Salcedo, E. Distributed Intelligent Video Surveillance for Early Armed Robbery Detection based on Deep Learning. *Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2024, pp. 297–314.
141. Han, Q.; Ren, X.; Zhao, P.; Wang, Y.; Wang, L.; Zhao, C.; Yang, X. ECCVideo: A Scalable Edge Cloud Collaborative Video Analysis System. *IEEE Intelligent Systems* **2023**, 38, 34–44. doi:10.1109/MIS.2022.3214614.
142. Chen, J.; Ran, X. Deep Learning With Edge Computing: A Review. *Proc. IEEE* **2019**, 107, 1655–1674.
143. Gorospe, J.; Mulero, R.; Arbelaitz, O.; Muguerza, J.; Antón, M.Á. A Generalization Performance Study Using Deep Learning Networks in Embedded Systems. *Sensors* **2021**, 21.
144. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size, 2016, [arXiv:cs.CV/1602.07360].
145. Fang, W.; Wang, L.; Ren, P. Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments. *IEEE Access* **2020**, 8, 1935–1944.
146. Karnin, E.D. A simple procedure for pruning back-propagation trained neural networks. *IEEE Trans. Neural Netw.* **1990**, 1, 239–242.
147. Yu, F.; Cui, L.; Wang, P.; Han, C.; Huang, R.; Huang, X. EasiEdge: A Novel Global Deep Neural Networks Pruning Method for Efficient Edge Computing. *IEEE Internet of Things Journal* **2021**, 8, 1259–1271.
148. Woo, Y.; Kim, D.; Jeong, J.; Ko, Y.W.; Lee, J.G. Zero-Keep Filter Pruning for Energy/Power Efficient Deep Neural Networks. *Electronics* **2021**, 10, 1238.
149. Zebin, T.; Scully, P.J.; Peek, N.; Casson, A.J.; Ozanyan, K.B. Design and Implementation of a Convolutional Neural Network on an Edge Computing Smartphone for Human Activity Recognition. *IEEE Access* **2019**, 7, 133509–133520.
150. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network, 2015, [arXiv:stat.ML/1503.02531].
151. Watrix. Watrix. WATRIX.ai, 2024. Accessed: Sept. 3, 2024.
152. NtechLab. NtechLab. NTECHLAB.com, 2024. Accessed: Sept. 3, 2024.
153. Ruiz-Barroso, P.; Castro, F.M.; Delgado-Escano, R.; Ramos-Cózar, J.; Guil, N. High performance inference of gait recognition models on embedded systems. *Sustainable Computing: Informatics and Systems* **2022**, 36, 100814. doi:10.1016/j.suscom.2022.100814.
154. Zeng, X.; Zhang, X.; Yang, S.; Shi, Z.; Chi, C. Gait-based implicit authentication using edge computing and deep learning for mobile devices. *Sensors* **2021**, 21, 4592.

155. Yoshino, K.; Nakashima, K.; Ahn, J.; Iwashita, Y.; Kurazume, R. Gait recognition using identity-aware adversarial data augmentation. 2022 IEEE/SICE International Symposium on System Integration (SII). IEEE, 2022, pp. 596–601.
156. Li, Z.; Li, Y.R.; Yu, S. FedGait: A Benchmark for Federated Gait Recognition. 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, 2022, pp. 1371–1377.
157. Das, D.; Agarwal, A.; Chattopadhyay, P. Gait recognition from occluded sequences in surveillance sites. European Conference on Computer Vision. Springer, 2022, pp. 703–719.
158. Sepas-Moghaddam, A.; Etemad, A. View-invariant gait recognition with attentive recurrent learning of partial representations. *IEEE Transactions on Biometrics, Behavior, and Identity Science* **2020**, 3, 124–137.
159. Hadid, A.; Ghahramani, M.; Kellokumpu, V.; Pietikäinen, M.; Bustard, J.; Nixon, M. Can gait biometrics be Spoofed? Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012, pp. 3280–3283.
160. Hadid, A.; Ghahramani, M.; Kellokumpu, V.; Feng, X.; Bustard, J.; Nixon, M. Gait biometrics under spoofing attacks: an experimental investigation. *Journal of Electronic Imaging* **2015**, 24, 63022. doi:10.1117/1.JEI.24.6.063022.
161. Jia, M.; Yang, H.; Huang, D.; Wang, Y. Attacking Gait Recognition Systems via Silhouette Guided GANs. Proceedings of the 27th ACM International Conference on Multimedia; Association for Computing Machinery: New York, NY, USA, 2019; MM '19, pp. 638–646.
162. Hirose, Y.; Nakamura, K.; Nitta, N.; Babaguchi, N. Discrimination between genuine and cloned gait silhouette videos via autoencoder-based training data generation. *IEICE Trans. Inf. Syst.* **2019**, E102D, 2535–2546.
163. Maqsood, M.; Ghazanfar, M.A.; Mehmood, I.; Hwang, E.; Rho, S. A Meta-Heuristic Optimization Based Less Imperceptible Adversarial Attack on Gait Based Surveillance Systems. *Journal of Signal Processing Systems* **2022**. doi:10.1007/s11265-022-01742-x.
164. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. *Advances in Engineering Software* **2014**, 69, 46–61. doi:https://doi.org/10.1016/j.advengsoft.2013.12.007.
165. Hirose, Y.; Nakamura, K.; Nitta, N.; Babaguchi, N. Anonymization of Human Gait in Video Based on Silhouette Deformation and Texture Transfer. *IEEE Transactions on Information Forensics and Security* **2022**, 17, 3375–3390. doi:10.1109/TIFS.2022.3206422.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.